VIA EFS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Patent Application of:	Bernhard KEPPLER		
Conf. No.:	2897	Group Art Unit:	1626
Appln. No.:	10/627,519	Examiner:	Rei Tsang Shiao
Filing Date:	July 25, 2003	Attorney Docket No.:	8182-25US (PA32424US)
Title:	COMPOSITIONS CONTAINING A RUTHENIUM(III) COMPLEX AND A HETEROCYCLE		

DECLARATION OF NIKOLAI GRAF V. KEYSERLINGK UNDER 37 C.F.R. §1.132

I, Nikolai Graf v. Keyserlingk, Ph.D., declare and state as follows:

1. I am currently a project manager of Faustus Forschungs Cie., Translational Cancer Research GmbH, the assignee of the above-referenced patent application.

2. I am a chemist with working experience in synthetic organic chemistry. In addition I have gained experience in the field of cell culture experiments, toxicology examination and clinical research during my work as project manager in the field of drug development for Faustus Forschungs Cie. Translational Cancer Research GmbH for several years. I believe that I would be recognized as a person at least of ordinary skill in the art to which the present invention pertains and I am familiar with the level of skill of such a person in this art. Please refer also to my *curriculum vitae* ("CV") attached hereto (**Attachment 1**).

3. I have reviewed the Examiner's Office Action dated January 5, 2006, in the above-referenced application, and I understand that the Examiner has rejected claim 37 under 35 U.S.C. §112, first paragraph, as lacking enablement. I understand that the Examiner contends that the Specification, while enabling for methods of treating colon cancer and ovarian cancer,

does not reasonably provide enablement for methods of using the presently claimed compositions for treating other types of cancer, without resort to undue experimentation.

4. I make this Declaration to correct what I view as a misconception on the part of the Examiner, upon which this enablement rejection appears based, and to demonstrate why the Specification provides enablement for the full scope and breadth of claim 37. Accordingly, below I explain how, based on the Specification and the knowledge of those skilled in the art, one can readily design the normal type of protocol and experimentation (which is not undue in the field of anti-tumor agents) which would be required to select a suitable dosage, regimen and/or route of administration for treating various types of tumors with a ruthenium (III) complex of the present invention. Additionally, I have described two evaluations that we have carried out to demonstrate both efficacy and to make initial determinations regarding dosage levels, dosage regimens and route of administration. I describe both a clinical human trial and an alternative *in vitro* method which can be used.

5. In the above-referenced application, claim 37 is directed to a method of inhibiting tumor activity by administering to a patient a composition according to claim 33. The detailed description set forth in the Specification, including preparation of the compositions, routes of administration and dosage levels, provides one of ordinary skill in the art with adequate information to enable such a person to make and use the invention as claimed. The amount of experimentation which may be required to determine an ideal dosage, regimen or route of administration for any particular type of tumor would not be undue. Moreover, one of ordinary skill in the art would necessarily expect such experimentation with respect to dosage levels and regimens, along with any particularly preferred route of administration to be required in each given instance.

6. The protocol of our Clinical Phase I Study (as detailed herein) was designed based on routinely gathered preclinical data obtained regarding efficacy and a toxicology evaluation, and was mainly designed to ensure the patients' safety. The primary objective of such a Phase I trial is establishment of the Maximum Tolerated Dose ("MTD") and

NL

safe starting doses for subsequent Phase II trials. Examination of efficacy is a secondary objective in a Phase I trial. For any person skilled in the art it is a well known fact that in Phase I trials, efficacy is not normally established from a regulatory perspective.

There are different options for the design of the schedule. In non-cancer indications it is standard to administer the study medication only one time in healthy individuals in the first clinical trial in order to get information mainly about pharmacokinetic behavior and tolerability in man. In cancer indications only patients suffering from cancer, who have experienced all possible standard therapies and for whom no other treatment options are available, are included in this first human trial (despite clinical efficacy not being a primary objective). Therefore a trial design which may hopefully increase the chance of providing the patient with a real benefit in terms of response of the tumor to the treatment is highly recommended, at least from an ethical point of view. Therefore, it was decided by the inventors, the investigators and the sponsor (Faustus) during the preparation of the protocol to administer the agent over three weeks, twice weekly, in order to maximize the chance seeing signs of efficacy without affecting the safety of the patients.

The starting dose was based on the MTD determined in the acute toxicity testing in animal studies carried out in accordance to international guidelines (CPMP/SWP/997/96, **Attachment 2**). In mice and rats, the data from the more sensitive species are used to establish the starting dose in the first human trial. The Clinical Phase I Study was designed as a dose escalation study using an accelerated dose escalation scheme (Simon et al., *J. Natl. Cancer Institute*, *89*, 1997, 1138-47, **Attachment 3**). In the absence of toxicity, the dose is doubled for each successive patient.

The evaluation of tumor response was based on the RECIST criteria by comparison of the baseline evaluation to the response evaluation after the treatment (Therasse et a., *J. Nat. Cancer. Inst.*, *92*, 2000, 205-216, **Attachment 4**). The method is summarized as follows:

Baseline Evaluation:

At baseline, tumor lesions were categorised as follows:

- measurable: lesions that could be accurately measured in at least one dimension (longest diameter to be recorded) as 20 mm with conventional techniques or as 10 mm with spiral CT scan;
- non-measurable: all other lesions, including small lesions (longest diameter < 20 mm with conventional techniques or < 10 mm with spiral CT scan) and;
- truly non-measurable lesions.

All measurable lesions up to a maximum of five lesions per organ and 10 lesions in total, representative of all involved organs, were identified as target lesions and recorded and measured at baseline. Target lesions were selected on the basis of their size (those with the longest diameter) and their suitability for accurate repeated measurements. A sum of the longest diameter for all target lesions was calculated and reported as the baseline sum longest diameter. This value was used as the reference by which to characterise the objective tumour response. All other lesions were identified as non-target lesions and were also recorded at baseline. Measurements of these lesions was not required, but the presence or absence of each was noted after the end of treatment.

Response Criteria:

The definition of criteria used to determine objective tumor response for target lesions was as follows, taking into account the measurement of the longest diameter only for all target lesions:

- Complete Response (CR): disappearance of all target lesions
- Partial Response (PR): at least a 30% decrease in the sum of the longest diameter of target lesions, taking as reference the baseline sum longest diameter
- Progressive Disease (PD): at least a 20% increase in the sum of the longest diameter of target lesions, taking as reference the smallest sum longest diameter of target lesion recorded since the treatment started or the appearance of one or more new lesions
- Stable Disease (SD): neither sufficient shrinkage to qualify for partial response nor sufficient increase to qualify for progressive disease, taking as

reference the smallest sum longest diameter of target lesion recorded since the treatment started

The definition of criteria used to determine objective tumor response for non target lesions was as follows:

- Complete Response (CR): the disappearance of all non-target lesions
- Incomplete Response/Stable Disease (SD): the persistence of one or more non-target lesions
- Progressive Disease (PD): the appearance of one or more new lesions and/or unequivocal progression of existing non-target lesions.

All these techniques and parameters are common state of the art and are a part of the normal practice and knowledge for a clinician/physician active in oncology.

7. Accordingly, as explained above in paragraph 6, we developed a protocol for initial investigation of dosage and treatment cycle for a variety of different tumors in human subjects (the "Clinical Phase I Study"). To further explain the normal (and not undue) experimentation that one of ordinary skill in the art could undertake in connection with applying the present invention to a variety of different tumors and in support of the broad efficacy of ruthenium complexes the present invention, the dosage experiments and clinical evaluations carried out by the physicians in the hospitals (two centers were recruiting patients under the supervision of the Faustus Clinical Research Department) in our Clinical Phase I Study are described below.

8. The data set forth below in Table 1 were obtained during the Clinical Phase I Study of a ruthenium (III) complex combination in accordance with the claimed invention. Such clinical studies can be readily designed and/or modified by one of ordinary skill in the art to ascertain efficacy of any given dosage, regimen and/or route of administration, based upon the knowledge of one of ordinary skill in the art, in conjunction with that which is disclosed in the Specification, for example, in Paragraph [0045] through Paragraph [0052] and in Examples 2 and 3.

9. In the Clinical Phase I Study, eight human subjects with different types of tumors were treated in an open dose-escalation-study with "Combination I." Combination I refers to a mixture of sodium-*trans*-[tetrachlorobis (1*H*-indazole) ruthenate (III)] and indazole-hydrochloride used in the study. The two components in Combination I were combined in a molar ratio of 1:1.1. Combination I was prepared by dissolving 50 mg of sodium-*trans*-[tetrachlorobis (1*H*-indazole) ruthenate (III)] in 15 ml of isotonic sodium chloride solution. The mixture was transferred to a sterile container and a volume of 116.5 ml of indazole-hydrochloride was added to the sterile container. After mixing, the mixed combination was used directly in the form of an infusion. The indazole-hydrochloride used comprised 16.95 mg of indazole-hydrochloride dissolved in a physiological salt solution. The patients received the medication twice a week over three consecutive weeks. Thus, one cycle of treatment corresponded to six administrations.

10. Each of the eight patients was a heavily pre-treated subject having already undergone treatment, including surgery, radiation and/or several courses of alternative chemotherapeutic agents. In each instance, the prior treatment failed to cure them or to keep them stable. The dosages and results of this study are summarized in Table 1 set forth below. The description of the results is in accordance with the internationally recognized RECIST Criteria explained above. Specifically, the results are categorized as:

CR = COMPLETE RESPONSE
PR = PARTIAL RESPONSE
SD = STABLE DISEASE
PD = PROGRESSIVE DISEASE

Table 1

Patient #	Diagnosis at Study Entry	Additional Cycles Yes/No	Dose (mg)	Results
1	Organ: sigmoid colon Histology: adenocarcinoma	Yes (2 cycles)	25	SD 9 weeks
2	Organ: rectum Histology: adenocarcinoma	No	50	Patient expired prior to completion of one cycle
3	Organ: colon Histology: adencarcinoma	No	50	SD 10 weeks
4	Bladder carcinoma	No	100	PD
5	Organ: liver Histology: cholangio cellular carcinoma	Yes (2 cycles)	200	SD 8 weeks
6	Organ: endometrium Histology: carcinoma	No	400	SD 10 weeks
7	Organ: left eye, melanoma of the choridea Histology: spindle bcell melanoma	No	600	PD
8	Site: tongue Histology: carcinoma	Yes (2 cycles)	600	SD 8 weeks

11. Table 1 and the Clinical Phase I Study carried out on the eight human subjects having various tumors show that one of ordinary skill in the art can easily design a protocol necessary for determining the dosage level appropriate for stabilizing and/or inhibiting further tumor growth. The Clinical Phase I Study was designed as a dose escalation trial and eight patients were treated at different dose levels. As can be seen from the Results data in table 1, responses to the treatment were observed over a comparable broad range of dosages. One rationale for performing such a Clinical Phase I Study is to increase the dose to be administered in order to determine the MTD in the clinical situation. Based on the obtained data in the Clinical Phase I Study further Phase II trials using the ruthenium(III) complex (*i.e.*, Combination I) will use 600 mg as starting dose.

The type of experimentation which may be necessary to establish efficacy and safety in accordance with the administration of such compounds is well within the normal efforts of one ordinary skill in the art and is not considered undue experimentation. According to our protocol, only patients for which no other treatment options were available were included in the study. These patients have received all standard therapies for their condition and every therapy

failed. All patients in the Study, including Patients 2, 4 and 7, belong to this category. Patient 2 died during the first treatment cycle due to progressive worsening of his already advanced condition. Inclusion of Patient 2 in the Study results is not in accordance to the protocol which requires sufficient life expectancy for the administration of at least one whole cycle of trial medication and subsequent observation of the patient. Although Patient 4 (bladder carcinoma) and Patient 7 (melanoma) did not respond to the administration of Combination I, it is not clearly ruled out that these kind of tumors are insensitive to the treatment with the study medication. In fact, it can be seen from Table 2 (below), that Combination I showed *in vitro* activity in a melanoma cell line. In the Clinical Phase I Study, individual aspects such as current patient performance, pharmacokinetic behaviour and/or prior treatments can play a role in determining tumor response.

12. In addition, the activity and efficacy of the inventive compositions in numerous tumor indications can be observed on an *in vitro* basis. Suitable dosage levels and regimens can be established, based upon such *in vitro* studies, for use in subsequent clinical evaluations. For example, the activity of Combination I was evaluated by measuring the growth of various carcinoma cell lines treated *in vitro* with Combination I, as described below. One skilled in the art is capable of planning and carrying out such an *in vitro* evaluation.

13. Cell culture experiments were performed according to known procedures described in literature (P. Skehan et al., *J. Nat. Canc. Inst.* 1990, *82*, 1107-1112, **Attachment 5**). The various cell lines were plated out for an assay of the activity of Combination I. Combination I and the positive control, Cisplatin, were dissolved in cell culture medium in serial dilutions ranging from $2x10^{-3}$ to $2 x10^{-7}$ mol/L. Plate 1 was used to establish the starting concentration of the cells (*i.e.*, reference value in order to compare the treated cells and untreated controls after finalization of the assay with the original starting concentration of the cells) . All other plates were used for the assay and were incubated either with medium (negative control cells), with Cisplatin (positive control cell) or the Combination I in varying concentrations (treated cells). Approximately 100 µl of unfixed and re-suspended cell suspension was added to each well and

MG

incubated at 37°C at 95% relative humidity and 5% carbon dioxide for 24 hours. Cells of plate 1 were fixed with 25 ml of trichloroacetic acid and after two hours at 4°C, washed with tap water and air dried. Approximately 100 µl of each serial dilution of Combination I, positive (Cisplatin) and negative (only Medium) controls were added to the cells in triplicate and incubated for 48 hours at 37°C, 95% relative humidity and 5% carbon dioxide. The cells of the assay plates were fixed with 50 µl of trichloroacetic acid and after two hours at 4°C, were washed with tap water and air dried. The number of cells was detected using sulforhodamine B, a dye which is capable of binding with cellular proteins. Cell number is detected by measuring the extinction which is proportional to the concentration of the dye and which represents the number of cells in the assay. Approximately 100 µl of 0.4% sulforhodamine B ("SRB") in 1% acetic acid were added to each well. The SRB was removed with the acetic acid and the plates were air dried. Approximately 100 µl of 10 µM Tris Base were added to each well and a reading at a wavelength of 515 nm was taken in order to determine the extinction caused by the dye. The percentage growth of each carcinoma cell line was calculated for each concentration as follows:

$[(Ti-Tz)/(C-Tz)] \times 100$ If $Ti \geq Tz$

$[(Ti-Tz)/Tz] \times 100$ If $Ti < Tz$

where Ti represents the extinction of the cells treated with the respective concentration of either Combination I or the positive control (average values used from three different wells for each concentration of each of Combination I and the positive control); Tz represents the extinction of the cells which were directly fixed after the seeding and which represents the starting number of cells; and C represents the extinction of the cells grown in the control experiments without being contacted by either Combination I or the positive control agent, Cisplatin. This kind of calculation allows the determination of the inhibition of cell growth. The obtained values of percentage growth for each concentration in comparison to control cells are plotted into a graphic representation. The GI50 (= the concentration of either Combination I or the positive control, where the growth of the treated cells is reduced to an amount of 50% compared to untreated control cells) could be directly determined from this graphic representation. The results are set forth in Table 2 below.

Table 2.

Cell Line	Indication	GI50 (μM)
A431 (ec)	Epidermoid carcinoma	2.00×10^{-4}
PC3 (pc)	Prostate carcinoma	2.50×10^{-4}
SW480 (cc)	Colon carcinoma	2.75×10^{-4}
A549 (lc)	Lung carcinoma	1.75×10^{-4}
SK-RC-47 (rc)	Renal cell carcinoma	2.25×10^{-4}
FEMXI (m)	Melanoma	2.50×10^{-4}

14. From Table 2 and the experiment described in paragraph 13, it can be seen that one of ordinary skill in the art can readily design an *in vitro* protocol which can assist in determining an appropriate dosage and regimen for inhibiting tumor growth. Such *in vitro* activities do not necessitate the identification of human subjects and can allow one of ordinary skill in the art to make an initial determination of the level of active ingredient necessary to produce tumor inhibiting activity.

15. In conclusion, I believe one of ordinary skill in the art, upon reviewing the Specification of this application, would understand and be equipped to make and use the invention, in accordance with claim 37, for a variety of types of tumors and would understand the steps and methods necessary for designing a protocol, to determine a dosage and regimen appropriate for inhibiting tumor activity as disclosed. Moreover, the broad variety of tumors which show inhibited growth upon treatment according to various embodiments of the present invention is clearly evidenced.

I declare further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further, that those statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

31.06.2006
Date

Nikolai Graf v. Keyserlingk, Ph.D.

NK

Anhang 1

CV Nikolai Graf Keyserlingk, PhD

born	19.12.1971
in	Bielefeld, Germany

Education:

1991	Matura in Bremerhaven, Germany
1991 – 1992	Army service
1992 – 1997	Education in Chemistry at the University of Oldenburg, Germany, Diploma
1998-2001	PhD Thesis in synthetic organic chemistry at the University of Oldenburg

Working experience:

1998-2001	Scientific Co-worker, University of Oldenburg
since Nov. 2001	Project Manager Drug Development at Faustus Forschungs Compagnie, Leipzig, Germany

Publications:

1. N. Graf von Keyserlingk, J. Martens, D. Ostendorf, W. Saak und M. Weidenbruch: Utilization of industrial waste materials. Part 20. Stereoselective cycloaddition of silylenes and disililene to an enantiomerically pure cyclic ketimine derived from an industrial waste material, *J. Chem. Soc. Perkin I* **2001**, 706-710.

2. N. Graf von Keyserlingk und J. Martens: Synthesis of optically active functionalised cyclic ketimines and their application in enantioselective catalysis, *Tetrahedron Asymmetry* **2001**, *12*, 2213-2222.

3. N. Graf von Keyserlingk und J. Martens: Highly diastereoselective addition of *N*-Boc-pyrrolidin-2-yllithium to optically active ketimines: Synthesis of enantiomerically pure 1,3-imidazolidin-2-ones and diamines, *Eur. J. Org. Chem.* **2002**, 301-308.

4. M. I. Burguete, E. García-Verdugo, M. J. Vicent, S. V. Luis, H. Pennemann, N. Graf von Keyserlingk und J. Martens: New Supported β-Amino Alcohols as Efficient Catalysts for the Enantioselective Addition of Diethylzinc to Benzaldehyde under Flow Conditions, *Org. Lett.* **2002,** *4* 3947-3950.

5. M. I. Burguete, M. Collado, E. García-Verdugo, M. J. Vicent, S. V. Luis, H. Pennemann, N. Graf von Keyserlingk und J. Martens: Development of small focused libraries of supported amino alcohols as an efficient strategy for the optimization of enantioselective heterogeneous catalysts for the $ZnEt_2$ addition to benzaldehyde, *Tetrahedron* **2003**, *59*, 1797-1804.

6. M. Galanski, C. Baumgartner, K. Meelich, V. B. Arion, M. Fremuth, M.A. Jakupec, P. Schluga, C. Hartinger, N. Graf v. Keyserlingk, B. K. Keppler,: Synthesis, crystal structure and pH dependent cytotoxicity of (SP-4-2)-bis(2-aminoethanolato-κ2N,O)platinum(II) – a representative of novel pH sensitive anticancer platinum complexes, *Inorganica Chimica Acta*, **2004**, 357, 3237-3245.

7. M. Galanski, A. Yasemi, M. A. Jakupec, N. Graf v. Keyserlingk and B. K. Keppler: Synthesis, Cytotoxicity, and Structure-Activity Relationships of New Oxaliplatin Derivatives, *Monatshefte für Chemie*, **2005**, *136*, 693-700.

8. L. Habala, M. Galanski, A. Yasemi, A. A. Nazarov, N. Graf v. Keyserlingk, B. K. Keppler: Synthesis and structure-activity relationships of mono- and dialkyl-substituted oxaliplatin derivatives, *Eur. J. Med. Chem.* **2005**, *40*, 1149-1155.

9. M. Galanski, L. Habala, A. A. Nazarov, M. A. Jakupec, A. Yasemi, S. Slaby, N. Graf von Keyserlingk, B. K. Keppler: Rational development of oxaliplatin analogues –synthesis and preliminary structure-activity relationships, *Int. J. Clin. Pharmacol. Ther.* **2005**, *43*, 575-6.

10. R. D. Hofheinz, C. Dittrich, M. A. Jakupec, A. Drescher, U. Jaehde, M. Gneist, N. Graf von Keyserlingk, B. K. Keppler, A. Hochhaus: Early results from a phase I study on orally administered tris(8-quinolinolato)gallium(III) (FFC11, KP46) in patients with solid tumors—a CESAR study (Central European Society for Anticancer Drug Research–EWIV), *Int. J. Clin. Pharmacol. Ther.*, **2005**, 43, 590-1.

11. M. A. Jakupec, V. B. Arion, S. Kapitza, E. Reisner, A. Eichinger, M. Pongratz, B. Marian, N. Graf von Keyserlingk, B. K. Keppler: KP1019 (FFC14A) from bench to bedside: preclinical and early clinical development– an overview, Int. J. Clin. Pharmacol. Ther. **2005**, *43*, 595-6.

12. P. Heffeter, M. A. Jakupec, W. Korner, S. Wild, N. Graf von Keyserlingk, L. Elbling, H. Zorbas, A. Korynevska, S. Knasmuller, H. Sutterluty, M. Micksche, B. K. Keppler, W. Berger: Anticancer activity of the lanthanum compound [tris(1,10-phenanthroline)lanthanum(III)] trithiocyanate (KP772; FFC24), *Biochem Pharmacol.* **2006**, 71, 426-40.

13. A. Dobrov, V.B. Arion, N. Kandler, W. Ginzinger, M. A. Jakupec, A. Rufinska, N. Graf von Keyserlingk, M. Galanski, C. Kowol, B. K. Keppler: The first metal-based paullone derivative with high antiproliferative activity in vitro, *Inorg Chem.* **2006**. 45, 1945-50.

Attachment 2

PC 6.7.1

The European Agency for the Evaluation of Medicinal Products
Human Medicines Evaluation Unit

London, 23 July 1998
CPMP/SWP/997/96

COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS (CPMP)

NOTE FOR GUIDANCE ON THE PRE-CLINICAL EVALUATION OF ANTICANCER MEDICINAL PRODUCTS

DISCUSSION IN THE SAFETY WORKING PARTY (SWP)	December 1996
TRANSMISSION TO THE CPMP	December 1996
RELEASE FOR CONSULTATION	December 1996
DEADLINE FOR COMMENTS	March 1997
DISCUSSION IN THE SAFETY WORKING PARTY	November 1997 June 1998
FINAL APPROVAL BY THE CPMP	July 1998
DATE FOR COMING INTO OPERATION	January 1999

7 Westferry Circus, Canary Wharf, London E14 4HB, UK
Tel: (+44-171) 418 84 00 Fax: (+44-171) 418 85 51
E_Mail: mail@emea.eudra.org http://www.eudra.org/emea.html

PRE-CLINICAL EVALUATION OF ANTICANCER MEDICINAL PRODUCTS

1. INTRODUCTION

1.1 Objectives of the guideline

The purpose of this guideline is to define the preclinical data which are considered obtainable from preclinical studies with respect to pharmacodynamic, pharmacokinetic and toxicological properties of new anticancer drugs and which are considered relevant with respect to Phase I (Human Pharmacology), Phase II (Therapeutic Exploratory) and Phase III (Therapeutic Confirmatory) Clinical Trials and Marketing Applications.

Furthermore, the guideline serves the purpose of avoiding unnecessary tests, thus enabling the promptest possible introduction of newly developed anticancer medicinal products into clinical trials without compromising safety.

This note for guidance should be read in the light of general requirements set be Council Directive 75/318 (EEC) as amended. The applicant should also refer to the Note for Guidance on Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals (CPMP/ICH/286/95).

1.2 Scope of the guideline

The guideline concerns primarily cytotoxic/cytostatic drugs that are presumed to have a direct effect on tumour cells. It focuses on the development of single drug treatment. To support the clinical development of combinations of anticancer drugs, preclinical testing to investigate pharmacodynamic, kinetic and toxicological interactions is encouraged.

The guideline is aimed at formulating recommendations for pharmacodynamic investigations and the requirements for toxicological studies prior to Phase I, II and III Clinical Trials as well as Marketing Applications. As appropriate, additional studies may be required based on the findings of preclinical and clinical studies.

2. CHARACTERISATION OF PRIMARY PHARMACODYNAMICS

Prior to Phase I studies, preliminary characterisation of the mechanism(s) of action, resistance, and schedule dependencies as well as anti-tumour activity *in vivo* should have been made. As appropriate, these properties should be further investigated in parallel with Phase II and III studies.

2.1 In vitro studies

The primary aims of the *in vitro* studies are to obtain mechanistic information about the test substance and to characterise the activity profile.

2.1.1 Activity profile and mechanism(s) of action

By determination of the activity of a new drug at different concentrations in an appropriately selected cell panel and identifying IC_{50} concentrations for each cell line, a drug-specific activity profile is obtained. By comparing the profile with that of standard drugs, the activity of the new drug can be classified as similar or unrelated.

If a specific target structure is indicated, cell lines expressing different levels of this structure should be studied, if possible.

The use of well-characterised cell lines as regards genotype and biochemistry is encouraged. The selected test panel should be justified and the following panel should be considered:

- cell lines with different proliferation rates
- cell lines with different growth characteristics (e.g. solid and haematological)
- cell lines expressing general drug sensitivity and general resistance
- cell lines with sub-lines expressing specific resistance pheno/geno types

The use of cell line panels such as those found in the NCI cell line screen, which are well characterised with respect to sensitivity to standard agents, genotype, biochemistry and 'molecular targets', could also be accepted.

2.1.2 Mechanism(s) of resistance

In parallel with the characterisation of the mechanism(s) of action, the corresponding profile with respect to possible mechanism(s) of resistance (e.g. overexpression of P-glyco-protein/multidrug resistance protein/glutathione, changes in topoisomerase I and II) can be obtained.

Observed resistance could be investigated for its circumvention by resistance modulating agents. Investigation of the possible induction of resistance by long-term exposure of cell lines to the new drug and further characterisation of mechanism(s) of resistance are encouraged.

Assessment in the cell test panel of the activity of standard drugs in parallel with that of the new drug is recommended for establishing the existence of possible cross-resistance.

2.1.3 Exposure time and cell-cycle dependency

AUC normalised time dependency of drug activity and studies of cell-cycle- dependency of a new drug are recommended as an aid for the selection of proper dosing schedules. Studies in proliferating as well as non-proliferating cells are encouraged.

2.1.4 Disease-specific activity

The activity profile may be further investigated in fresh tumour samples from patients representing different diagnostic groups utilising justified techniques.

2.2 *In vivo* studies

The primary aims of *in vivo* studies are to obtain further information with respect to antitumour activity, therapeutic index and schedule dependency.

Studies in animals are usually carried out in rodents, mainly in mice, giving due consideration, when possible, to likely differences to man in pharmacokinetics/dynamics. The selection of a suitable animal model (including species, strain and tumour type) depends on the properties and proposed therapeutic indications of the anticancer drug and the available information about the response of different tumour cell lines. Anticancer drugs may be tested against xenografts of human cell lines inoculated in immunodeficient mice or tumour cell lines implanted in immunocompetent rodents. The type of tumour cell studied, the tumour load and the progression of the disease (e.g. metastases) in the animal should be considered.

The administration route and dosing regimen should mimic the anticipated clinical treatment schedule as far as possible.

Suitable criteria for the evaluation of efficacy include tumour growth, survival time and degree of remission or cure.

3. EVALUATION OF TOXICITY

The primary aims of the toxicity studies are to

- establish the maximal tolerated dose (MTD based on approximate minimal lethal dose) to be used to define the starting dose in Phase I trials (cf. section 3.3).
- identify effects on vital functions and target organ toxicity in relation to drug exposure and "treatment cycles" to support dose escalation in Phase I studies and duration of therapy.

3.1 Safety pharmacology

For compounds with a novel mechanism of action, an evaluation of safety pharmacology data (e.g. respiratory and cardiovascular effects) should have been made before the initiation of Phase I trials.

3.2 Pharmacokinetic/toxicokinetic studies

The evaluation of limited kinetic parameters, e.g. peak plasma levels and AUC, at doses around the MTD in the animal species used for preclinical studies may facilitate dose escalation during Phase I studies. Further information on ADME in animals should normally be made available prior to Phase II/III studies.

3.3 Single-dose toxicity studies

An assessment of those dose levels at which severe toxic symptoms or death occur (limit dose approach) should be performed in rodents with the administration route and formulation envisaged for clinical use.

A preliminary dose-finding study should be performed to establish an approximate MTD (maximal dose compatible with survival) in mice followed by a study with additional doses and animals to establish the MTD more accurately. The findings should be confirmed in rats to establish whether the relationship between toxicity and surface area is linear. If not, the Phase I starting dose should be based on the most sensitive species.

Dosages and the required number of animals per dose should be determined on the basis of the previous results in such a way that the necessary accuracy will be achieved with a minimum number of animals. The follow-up period of observation for the surviving animals should be at least 14 days.

The MTD, as established from single-dose toxicity studies, should be known prior to Phase I trials. Experience has shown that one tenth of the MTD may be an appropriate starting dose in Phase I studies.

In cases where the rodent species are known to be poor predictors of toxicity in humans e.g. antifolates, or the agent under investigation has a novel mechanism of action, an approximate MTD should be established in a non-rodent species.

3.4 Repeat dose toxicity studies

The dosing schedule should be as similar to the proposed clinical schedule as possible. Particular attention should be paid to critical target organ toxicity and reversibility of toxic effects.

A repeat-dose toxicity study of limited duration (2 to 4 weeks or 1 to 2 cycles) in two rodent species should be performed prior to Phase I studies. For compounds with a novel mechanism of action studies should be performed in a rodent and a non-rodent species.

For Phase II and Phase III trials and for Marketing Applications, repeat-dose toxicity studies should be performed in a rodent and a non-rodent species. Irrespective of daily or intermittent administration in the clinic, the duration of the repeat dose toxicity studies should be at least equal to the duration of the clinical trials, although not longer than 6 months.

3.5 Genotoxicity/Carcinogenicity

Normally, there is no established therapy available for patients eligible for Phase I and II Trials Therefore, prior to Phase I and II Trials, genotoxicity testing is not required. *In vitro* genotoxicity tests should have been performed prior to Phase III trials and Marketing Application. Normally, carcinogenicity studies are not required (cf. ICH S1A).

3.6 Toxicity to Reproduction

Studies of toxicity to reproduction are not required since cytotoxic/cytostatic drugs are assumed to cause reproductive disturbances. Pregnant women may nevertheless be treated with these agents and therefore studies elucidating the potential for reproductive toxicity are encouraged.

3.7 Local tolerance

Anticancer drugs can be highly toxic to tissues which come into contact with the product. Prior to Phase I studies, an evaluation of local tolerance relevant to the intended route(s) of clinical administration and user safety of the investigational product should be made. It should be noted that local tolerance testing may be part of other toxicity studies provided that the product is given via the intended clinical route of administration. If the product intended for marketing differs from the investigational product, relevant local tolerance, including paravenous, should be considered prior to Phase III studies and Marketing Applications.

AHadmnt 3

Accelerated Titration Designs for Phase I Clinical Trials in Oncology

*Richard Simon, Boris Freidlin, Larry Rubinstein, Susan G. Arbuck, Jerry Collins, Michaele C. Christian**

Background: **Many cancer patients in phase I clinical trials are treated at doses of chemotherapeutic agents that are below the biologically active level, thus reducing their chances for therapeutic benefit. Current phase I trials often take a long time to complete and provide little information about interpatient variability or cumulative toxicity.** ***Purpose:*** **Our objective was to develop alternative designs for phase I trials so that fewer patients are treated at subtherapeutic dose levels, trials are of reduced duration, and important information (i.e., cumulative toxicity and maximum tolerated dose) needed to plan phase II trials is obtained.** ***Methods:*** **We fit a stochastic model to data from 20 phase I trials involving the study of nine different drugs. We then simulated new data from the model with the parameters estimated from the actual trials and evaluated the performance of alternative phase I designs on this simulated data. Four designs were evaluated. Design 1 was a conventional design (similar to the commonly used modified Fibonacci method) using cohorts of three to six patients, with 40% dose-step increments and no intrapatient dose escalation. Designs 2 through 4 included only one patient per cohort until one patient experienced dose-limiting toxic effects or two patients experienced grade 2 toxic effects (during their first course of treatment for designs 2 and 3 or during any course of treatment for design 4). Designs 3 and 4 used 100% dose steps during this initial accelerated phase. After the initial accelerated phase, designs 2 through 4 resorted to standard cohorts of three to six patients, with 40% dose-step increments. Designs 2 through 4 used intrapatient dose escalation if the worst toxicity is grade 0-1 in the previous course for that patient.** ***Results:*** **Only three of the actual trials demonstrated cumulative toxic effects of the chemotherapeutic agents in patients. The average number of patients required for a phase I trial was reduced from 39.9 for design 1 to 24.4, 20.7, and 21.2 for designs 2, 3, and 4, respectively. The average number of patients who would be expected to have grade 0-1 toxicity as their worst toxicity over three cycles of treatment is 23.3 for design 1, but only 7.9, 3.9, and 4.8 for designs 2, 3, and 4, respectively. The average number of patients with grade 3 toxicity as their worst toxicity increases from 5.5 for design 1 to 6.2, 6.8, and 6.2 for designs 2, 3, and 4, respectively. The average number of patients with grade 4 toxicity as their worst toxicity increases from 1.9 for design 1 to 3.0, 4.3, and 3.2 for designs 2, 3, and 4, respectively.** ***Conclusion:*** **Accelerated titration (i.e., rapid intrapatient drug dose escalation) designs appear to effectively reduce the number of patients who are undertreated, speed the completion of phase I trials, and provide a substantial increase in the information obtained. [J Natl Cancer Inst 1997;89:1138-47]**

There has been considerable recent interest in new designs for phase I clinical trials. With currently used designs, many patients are treated at doses below the biologically active level, minimizing the opportunity for antitumor response (*1*). Although most patients who participate in phase I trials hope to obtain

**Affiliations of authors:* R. Simon, L. Rubinstein, S. G. Arbuck, M. C. Christian, Cancer Therapy Evaluation Program, Division of Cancer Treatment, Diagnosis, and Centers, National Cancer Institute, Bethesda, MD; B. Freidlin, The Emmes Corporation, Potomac, MD; J. Collins, U.S. Food and Drug Administration, Rockville, MD.

Correspondence to: Richard Simon, D.Sc., National Institutes of Health, Executive Plaza North, Rm. 739, Bethesda, MD 20892.

See "Notes" following "References."

therapeutic benefit from promising new experimental treatments, few achieve this objective (*2*). Whereas most patients would not have derived benefit from drugs studied in phase I trials, even if treated at the maximum tolerated dose (MTD), treating patients at substantially lower doses is likely to further reduce whatever chance for benefit might exist.

A second problem with current designs is that phase I trials may take a long time to complete, especially when the starting dose is far below the MTD (*3*). Current phase I trials also provide almost no information about variability among patients in the dose that can be tolerated without dose-limiting toxicity (DLT) or about whether there is evidence of cumulative toxicity.

In phase I trials of new drugs, the starting dose is usually one tenth of the LD_{10} (i.e., the dose that is lethal to 10% of animals) in the most sensitive animal species in which toxicology studies have been performed. Dose steps are defined by a modified Fibonacci series in which the increments of dose for succeeding levels are 100%, 67%, 50%, and 40%, followed by 33% for all subsequent levels. Three patients are usually treated at a dose level and observed for acute toxicity for one course of treatment before any more patients are entered. If none of the three patients experience DLT, then the next cohort of three patients is treated at the next higher dose. If two or more of the three patients experience DLT, then three more patients are treated at the next lower dose unless six patients have already been treated at that dose. If one of three patients treated at a dose experiences DLT, then three more patients are treated at that same level. If the incidence of DLT among those six patients is one in six, then the next cohort is treated at the next higher dose. In general, if two or more of the six patients treated at a dose level experience DLT, then the MTD is considered to have been exceeded, and three more patients are treated at the next lower dose as described above. The MTD is defined as the highest dose studied for which the incidence of DLT was less than 33%. Usually dose escalation for subsequent courses in the same patient, intrapatient dose escalation, is not permitted.

In this article we will describe alternative phase I designs that attempt to overcome some of the problems described above. We will then report the results of a computer simulation study conducted to evaluate the performance of alternative designs. The designs will be evaluated with regard to safety, the extent to which they provide patients the opportunity to be treated at higher doses more likely to provide antitumor response, the number of patients and time required to complete the trial, and the amount of information obtained.

Several alternative approaches to the design of phase I trials have been discussed in previous years. Collins et al. (*3*) recommended accelerating the dose escalations in humans by using the plasma drug $C \times T$ (i.e., the area under the concentration versus time curve) value at the LD_{10} in the mouse as the target exposure. This provides a pharmacokinetic basis for dose escalation, but is limited to clinical situations where a sensitive assay for the active drug moieties is available and where interspecies pharmacodynamic differences do not exist for the drug.

Storer (*4*) introduced the concept that the objective of a phase I trial is to estimate the dose that causes DLT in a specified proportion (e.g., 25% of the patients), and that this MTD should be estimated by fitting a logistic model to the dose versus DLT data. Storer also proposed using a single patient per dose level until the first DLT is observed.

Several investigators (*5-10*) have considered Bayesian designs. This approach makes use of a model relating dose administered to the probability of DLT. The parameters of the model are unknown initially, but some prior probability distribution for their values is assumed to be available based on preclinical data or experience with other drugs. As patients are treated, the probability estimates of the unknown parameters are updated based on the actual toxicity experience observed. Each patient is assigned the dose predicted to result in DLT for a target percentage (e.g., 25%) of the population.

Mick and Ratain (*11*) used a linear model relating white blood cell (WBC) count nadir to dose and pretreatment WBC count. They sequentially estimated the regression parameters of the model as data accumulated and individualized the dose based on pretreatment WBC count in an attempt to achieve a specified optimal WBC count nadir. Their approach predicts the optimal dose for each patient is based on pretreatment patient characteristics.

None of the designs described above considers how patients should be treated after the first course, nor do they use information obtained from subsequent courses. Except for the approach of Mick and Ratain (*11*), they do not consider interpatient variability or use information about toxic effects less than DLT.

Sheiner et al. (*12,13*) have argued for the use of titration (or intrapatient dose-escalation designs) for evaluating drug efficacy for diseases where the condition of a patient remains stable over a period of time. Titration designs involve dose escalation within patients until the desired biologic effect is obtained. If analyzed properly, they can provide information about interpatient variability in dose–response effects. The analysis of titration designs has been studied (*14,15*), but this approach has not been discussed in the context of phase I trials in oncology.

Methods

Phase I Designs Studied

The designs we evaluated differ with regard to the escalation/de-escalation rules for the first-course treatment of subsequent cohorts of patients as indicated in the "Appendix" section. Design 1 is the standard design described above. The other dose-escalation methods are based on a four-grade scale for defining the highest level of overall toxicity during each course of therapy. This scale can be defined differently to accommodate different clinical situations. For the purposes of this article, we have related the toxicity experience to grading scales commonly used in oncology, such as the National Cancer Institute Common Toxicity Criteria, and have described the levels as follows: none-mild (grades 0-1), moderate (grade 2), dose limiting (grade 3), and unacceptable (grades 4-5). Consistent with recent practice, we have not considered grade 3 neutropenia unaccompanied by either fever or infection to be dose limiting. We have grouped no toxicity with grade 1 toxicity because of the difficulty of determining whether mild abnormalities are drug or illness related in patients with cancer.

Design 2 treats one patient per dose level until one patient exhibits DLT or two patients exhibit grade 2 toxicity during their first course of treatment. At that time, the escalation plan switches to design 1. That is, two additional patients are accrued at the dose that triggered the switch, and three to six patients are treated in that and each subsequent cohort. This approach offers the possibility of speeding up the trial and reducing the number of patients assigned to low doses. It uses the first instance of first-course DLT to trigger the switch as proposed by Storer (*4*). It also uses first-course grade 2 toxicity to provide an added element of caution. We use the second instance of grade 2 toxicity for practical reasons, since it is often difficult to determine whether a grade 2 toxicity is drug related in a heterogeneous population of very ill patients.

Designs 3 and 4 also use only one patient per cohort during the early stage of the trial, but they incorporate more rapid dose escalation by using double-dose steps during this stage. With design 3, the single-patient-cohort stage of the trial also terminates when one patient experiences first-course DLT or two patients experience first-course grade 2 toxicity. With design 4, this accelerated stage terminates when the first instance of DLT or the second instance of grade 2 toxicity is observed in any course of treatment. In either case, after the rapid escalation stage terminates, subsequent cohort sizes are three to six patients and single-dose escalation steps are used as in design 1.

The Appendix also describes two approaches to individualizing dose through intrapatient dose modification. Intrapatient modification option A is the one most commonly used. There is no intrapatient dose escalation, only de-escalation. If the toxicity is dose limiting or worse in a course of chemotherapy, then the dose is reduced one level for the next course. Otherwise, the dose stays the same for the next course. Intrapatient modification option B permits escalation for each patient if the toxicity is grade 0-1 in the previous course for that patient. If the toxicity is moderate (grade 2), the dose remains unchanged. However, if the toxicity is DLT or worse, the dose is reduced. Designs 3 and 4 use two-dose-step (100%) intrapatient escalations during the initial accelerated phase of the trial, although de-escalations are always by single-dose steps. We have combined the standard cohort escalation design with the standard intrapatient dose modification option (A) as design 1 and have combined accelerated cohort escalations with the intrapatient dose escalation option (B) as designs 2 through 4. We will also provide results, however, for the mixed designs such as escalation option A with designs 2 through 4.

The accelerated designs are intended for use in phase I trials of drugs that have not been used previously in humans, where only preclinical information will be available for selecting a starting dose. Starting doses in these cases are often quite low, and designs that limit the number of patients treated at very low doses may be particularly useful.

Methodologic Approach

To evaluate alternative phase I designs, we wished to use data from actual phase I trials as much as possible. This could not be done directly because past trials were conducted with a particular escalation plan and we wished to evaluate new plans. Instead, we fit a stochastic model to data from past phase I trials. We then simulated new data from the model with the parameters estimated from the actual trials and evaluated the performance of alternative escalation designs on these simulated data. For any particular phase I trial, we generated 1000 simulated sets of data to reliably estimate the relative performance of the alternative designs. We repeated this for 20 different actual phase I trials of nine different drugs.

We required that the model we used be able to represent different levels of worst toxicity, not just presence or absence of DLT, and that the toxicity level experienced in a particular course would be determined by the dose administered in that course and the total dose administered in previous courses. We required that both interpatient and intrapatient variability be represented. We used the following model. Suppose that the ith patient receives dose d_{ij} during course j and has received a total dose of D_{ij} for courses previous to j. We let the coefficient α represent the influence of prior total dose ($\alpha = 0$ indicates no cumulative toxicity) and let the magnitude of toxicity increase logarithmically with dose. We introduced a random number β_i, normally distributed with mean μ_β and variance σ_β^2. This variable represents the interpatient variability in sensitivity to the toxic effects of the drug. We also introduced a random number ε_{ij}, normally distributed with mean zero and variance σ_ε^2, to represent the intrapatient variability in toxic response for a given patient receiving a given dose. These terms and random variables determine the magnitude of worst toxicity represented by

$$y_{ij} = \log(d_{ij} + \alpha D_{ij}) + \beta_i + \varepsilon_{ij}. \qquad [1]$$

If this value y_{ij} was less than a specified constant K_1, then patient i was considered to have experienced less than grade 2 toxicity during course j with dose d_{ij}. If the value of y_{ij} was greater than K_1 but less than K_2, then the toxicity level was taken to be grade 2; if the value was greater than K_2 but less than K_3, then the toxicity was considered to be dose limiting; and if y_{ij} was greater than K_3, then the toxicity was considered unacceptable. The values of the random numbers β_i vary across patients, but the same β_i was used for all treatment courses of the ith patient, while the within patient variability values ε_{ij} change from patient to patient as well as across courses.

This model can be viewed as a generalization of the $K_{\max}$ model used by Sheiner et al. (*12,13*). The above expression is equivalent to

$$\frac{e^{y_{ij}}}{1+e^{y_{ij}}} = \frac{(d_{ij}+\alpha D_{ij})}{e^{-(\beta_i+\varepsilon_{ij})}+(d_{ij}+\alpha D_{ij})}. \qquad [2]$$

The right-hand side of this equation is similar to the $K_{\max}$ model. The stimulus is of the form $d_{ij} + \alpha D_{ij}$ and the level giving 50% maximum response (exclusive of cumulative toxicity) is taken as a random variable, with mean approximately $e^{-\mu_\beta}$ and with a component identified with interpatient variability and a component associated with intrapatient variability. Our model measures toxicity in a categorical rather than continuous manner. Since the scale of the constants K_1, K_2, and K_3 is arbitrary, the fact that the left side of the equation involves a transformation of the originally defined y_{ij} does not matter. In fact, it can be shown that our model can be viewed as a generalization of the model of Chou and Talalay (*16*) in which the stimulus $d_{ij} + \alpha D_{ij}$ and 50% value are raised to a power p. With a categorical response in which the K's may be fit from the data, however, the power p is not identifiable, and the model is equivalent to that shown in equation 1.

The value σ_ε^2 represents the amount of intrapatient variability unexplained by current and previous doses. Setting $\sigma_\varepsilon^2 = 0$ means that the toxicity experienced by a patient is determined entirely by the doses and by patient characteristics that do not change from day to day. The value of σ_β^2 represents the amount of interpatient variability. Setting $\sigma_\beta^2 = 0$ means that patients entered in the clinical trial do not differ in their ability to tolerate the drug under study.

For these simulations, we used 40% increments between dose levels. With 40% increments, two-dose levels represent approximately a doubling of the dose because $1.4^2 = 1.96$. A 40% increment is close to the 33% increment that is used after the first few dose levels of trials based on the modified Fibonacci approach with which phase I investigators are familiar. Because interpatient variability in patient pharmacokinetic parameters and intrapatient variability in day-to-day susceptibility to toxicity are often substantial, it is usually not realistic to expect that one can estimate more precisely than to within 40% the dose that will give a desired level of biologic effect (*17*).

For all the simulations, we used $\mu_\beta = 0$, although the results are independent of this parameter. Table 1 shows the maximum likelihood estimates of the model parameters for the 20 actual phase I clinical trials studied. These trials were selected for a related study of nonstandard dose-escalation procedures. Although they were selected initially because they were planned to use nonstandard dose-escalation methods, only 9.5% of the patients received intrapatient dose escalation. Detailed information about the characteristics of these trials will be addressed in a separate report.

Only three of the 20 trials showed any evidence of substantial cumulative toxic effects as seen from the column labeled α in Table 1. Two of these studies involved the drug pyrazine diazohydroxide (PZDH) administered as a bolus every 3 weeks initially, but the interval between courses was eventually lengthened to 4-6 weeks because of delayed recovery from myelosuppression. Trial T90-156 administered PZDH daily for 5 days every 4-6 weeks, and no evidence of cumulative toxicity was obtained from our model parameters for that trial. The third trial showing evidence of cumulative toxicity involved flavone acetic acid (FAA). This latter trial was the only phase I trial with FAA that demonstrated cumulative toxicity. It differed from the other four FAA trials in that it used a weekly schedule of administration.

The standard deviations for intrapatient (σ_ε) and interpatient (σ_β) variability varied substantially. The larger values of σ_ε seen could represent true biologic variability or may reflect the difficulty of distinguishing drug-related toxicity from manifestations of illness for very sick patients. We used the original treating physician's assessment as to whether toxicity was drug related. Many of these patients were taking concomitant medications (not anticancer drugs) that may have influenced the toxicity experienced, and, in some cases, there may also have been nonstandardized treatment delays as a result of previous toxicity. With prospective use of titration designs, we expect that there will be more attention to these issues than could be the case in a retrospective analysis of a database.

The K_i value is given in terms of (K_i – log starting dose)/log 1.4 because this value represents approximately the number of 40% dose steps between the starting dose and the dose at which the average patient has a 50% chance of experiencing grade 2 or worse toxicity (since $\mu_\beta = 0$). The distance between other K values is similarly presented. Seven of the actual trials did not have any patients who experienced grade 4 toxicity. For these cases, the estimate of K_4 is very large by default, but the specific value is not meaningful.

Table 1. Estimates of model parameters for 20 phase I clinical trials

Drug	Trial	α	$(K_1 - \ln d_0)/\ln 1.4$*	$(K_2 - K_1)/\ln 1.4$	$(K_3 - K_2)/\ln 1.4$	σ_β	σ_ε
Flavone acetic acid	85-168	0	16.2	6.9	35†	.26	1.9
Flavone acetic acid	85-244	0	16.1	8.4	29†	2.9	.85
Flavone acetic acid	86-004	0	4.4	2.4	0.95	.47	.59
Flavone acetic acid	86-017	.24	8.0	2.9	2.2	0	.83
Flavone acetic acid	86-060	0	18.5	6.4	20†	.006	2.8
Piroxantrone	86-227	.08	8.4	2.7	2.3	1.03	.42
Piroxantrone	86-268	0	16.4	13.3‡	9.5‡	0	1.8
Chloroquinoxaline sulfonamide	88-114	.04	17.3	2.6	1.6	.88	.87
Chloroquinoxaline sulfonamide	88-127	0	13.7	4.6	2.9	.62	.90
Pyrazine diazohydroxide	89-053	.56	6.7	1.3	2.0	.37	.50
Pyrazine diazohydroxide	89-175	.24	6.6	1.3	0.53	.002	.65
Pyrazine diazohydroxide	90-156	.02	4.6	.53	.56	.001	.18
Pyrazoloacridine	90-073	.04	8.9	1.0	1.3	.24	.32
Cyclopentenylcytosine	91-018	0	4.4	.83	0.18	.19	.26
Fostriecin	91-106	.04	3.5	3.6	4.5	1.06	.54
Fostriecin	91-196	0	6.3	7.2	18†	.58	1.6
9-Aminocamptothecin	92-108	0	6.4	.48	0.39	.24	.11
9-Aminocamptothecin	92-186	0	6.0	.51	1.1	.35	.27
Penclomedine	93-087	.05	6.0	3.7	15†	.68	.81
Penclomedine	93-125	0	5.8	2.0	17†	.43	.53

*d_0 = starting dose.
†No grade 4 toxicity.
‡No grade 3+ toxicity.

It may be noted in Table 1 that the parameter estimates for different trials of the same drug sometimes vary substantially. This is due to a variety of causes, but the estimates provide a wide range of conditions for generating simulated data with which to compare alternative escalation designs.

Results

Comparison of Designs

The distribution of the highest dose level at which fewer than two instances of DLT occurred was very similar for the four designs for all of the 20 sets of parameters studied (Fig. 1). The true MTD was defined as the largest dose level for which the probability of first-course DLT or worse was less than .25, computed from the model using each set of parameters in Table 1. For simulations with each set of parameters, we tabulated the accuracy of the highest dose level with fewer than two instances of first-course DLT as a predictor of the true MTD. Fig. 1 shows that the four designs performed similarly in this regard. Although designs 2 through 4 use many fewer patients than design 1, in the dose range of interest, they have similar sample sizes. As will be seen later, fitting the model to data from a phase I trial provides a much richer set of information with which to plan phase II development. Fig. 1 demonstrates, however, that even with regard to the traditional estimate of phase II dose, accuracy is not sacrificed by the accelerated designs.

Fig. 2 shows histograms of the average number of patients required in the simulated trials for each design. In each graph, the sum of the heights of the bars is 20, the number of sets of parameters that is simulated. The x axis represents the average number of patients accrued in the 1000 simulated trials with each of the 20 sets of parameters. The standard design has a very broad distribution of sample size. For six of the sets of parameters, design 1 required more than 55 patients. For the 20 sets of parameters, design 1 required an average of 39.9 patients (median, 36.7 patients). Design 1 required substantially more patients than did the other designs. Design 2, which uses single 40% dose steps, required an average of only 24.4 patients (median, 21.8 patients). As seen in Fig. 2, the distribution of the number of patients is much narrower for design 2 than for design 1. Designs 3 and 4 also compare very favorably with design 1 with mean numbers of patients of 20.7 and 21.2, respectively, and median numbers of patients of 19.3 and 19.1, respectively. Design 2 does not require many more patients than the accelerated designs that use double dose steps. As will be seen below,

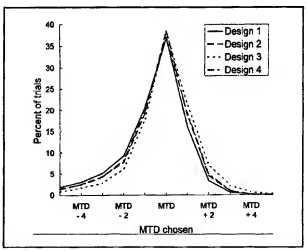

Fig. 1. Distribution of the maximum tolerated dose (MTD) chosen in simulated phase I trials averaged over the 20 sets of model parameters and 1000 replications for each set of parameters. In simulations, MTD is chosen in the traditional way as the largest dose level at which six patients are started and fewer than two experience first-course dose-limiting toxicity (DLT) or worse. True MTD is defined as the highest dose level for which the probability of first-course DLT or worse is less than .25, computed from the model with the use of each set of parameters in Table 1.

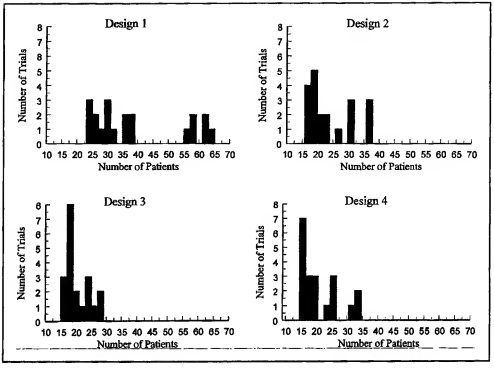

Fig. 2. Histograms of the average numbers of patients for simulated phase I trials with each of the 20 sets of model parameters. Averages are based on 1000 replications for each set of parameters. Total height of bars in each panel equals 20, the number of sets of model parameters.

the important difference between design 1 and the others is largely due to a reduction in patients treated early at subtherapeutic doses, where designs 2 through 4 accrue only one patient per level.

Another question of some importance is whether a reduction in the number of patients translates into a reduction in the duration of the trial. When eligible patients are very limited, the number of patients is closely associated with the duration of the trial. But if eligible patients are readily available, then it would take little more time to place three patients on a dose level than to place a single patient. Therefore, we also tabulated the number of cohorts required for each design, as shown in Fig. 3. The advantage of design 2 over design 1 with regard to the average number of patients does not translate into an advantage in the number of cohorts required. In fact, design 1 requires slightly fewer cohorts because design 2 sometimes overshoots its target and requires more cohorts at de-escalated levels. Designs 3 and 4, however, show substantial savings over designs 1 and 2 because of their use of double dose steps during the initial stage of the trials.

Fig. 4 shows the toxicity experience in the application of these designs to the phase I trials. In these simulations, we have assumed that all patients stay in the study for three courses of treatment and have tabulated the distribution of worst toxicity over these courses for each patient. For each set of parameters and each design, we have calculated the average number of patients whose worst toxicity was grade 0-1, grade 2, grade 3, or grade 4. This average was computed based on 1000 simulations for each of the 20 sets of design parameters. With the standard design 1, the average number of patients who have grades 0-1 toxicity as their worst toxicity over three cycles of treatment is 23.3. This number is substantially reduced for all of the newer designs; 7.9 for design 2, 3.9 for design 3, and 4.8 for design 4. Therefore, the number of undertreated patients is substantially reduced. This reduction is achieved with some increase in the number of patients with worst toxicity grade 3 or 4. The average number of patients with worst toxicity grade 3 increases from 5.5 with design 1 to 6.2, 6.8, and 6.2 for designs 2, 3, and 4, respectively.

Fig. 4 shows that the average number of patients with grade 4 toxicity increased from 1.9 with design 1 to 3.0, 4.3, and 3.2 for designs 2, 3, and 4, respectively. Hence, in comparing design 2 to design 1, on average, there is a reduction of about 15 patients per trial whose highest level of toxicity is grade 0-1 and an average increase of 1.8 patients per trial whose highest level of toxicity is grade 3-4. Design 4 provides a reduction of about 18 undertreated patients per trial and an average increase of about 2.1 overtreated patients. Design 3 provides a reduction of about 19 undertreated patients per trial, for an average increase of 3.7 overtreated patients. Hence, design 3 appears to have no real advantage over design 4. Although the average number of patients with worst toxicity grade 3-4 is not substantially increased using designs 2 through 4 compared with design 1, the proportion of patients with grades 3-4 toxicity is substantially increased. This is because designs 2 through 4 substantially reduce the expected number of patients with worst toxicity grade 0-1 and the total number of patients on trial compared with design 1. With design 1, a weighted average (taken over the 20 parameter sets, weighted by average sample size) of about 18% of patients experience grade 3-4 toxicity during some course of treatment. For designs 2, 3, and 4, the percentages are about 38%, 53%, and 45%, respectively. For grade 4 toxicity alone, the

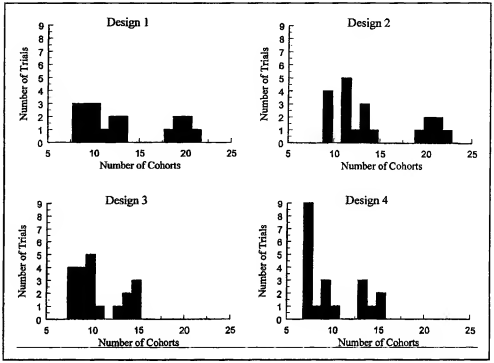

Fig. 3. Histograms of the average number of cohorts of patients for simulated phase I trials with each of the 20 sets of model parameters. Averages are based on 1000 replications for each set of parameters. Total height of bars in each panel equals 20, the number of sets of model parameters. Number of cohorts reflects time to completion when there is an excess of patients available for entry in the trial.

percentages are 5%, 12%, 20%, and 15% for designs 1 through 4, respectively.

There are six sets of parameters by use of design 1 for which three or more patients are expected to experience grade 4 toxicity. The trial with the largest number of such patients was T89-175. This is a PZDH trial with $\alpha = .24$. The PZDH trial with $\alpha = .56$ (T89-053) and the FAA trial with $\alpha = .24$ are also included in this set of six trials. It is not surprising that trials with a substantial amount of cumulative toxicity should result in patients experiencing grade 4 toxicity, even without using intrapatient dose escalation. The other three trials for which there were three or more patients expected to experience grade 4 toxicity using design 1 were T86-004, T88-114, and T91-018. These three trials are characterized by a combination of very steep dose–toxicity curves [as indicated by small values of $(K_3-K_2)/\ln(1.4)$] and relatively large amounts of intrapatient variability. With designs 2 or 4, the increase in the expected number of grade 4 toxic effects compared with design 1 is one patient or fewer in 12 of the 20 trials. The increase is greater than three patients in the three trials (T90-156, T91-018, and T92-108) characterized by very steep dose–toxicity curves. The increase in incidence of grade 4 toxicity was greater for design 3 than for designs 2 or 4.

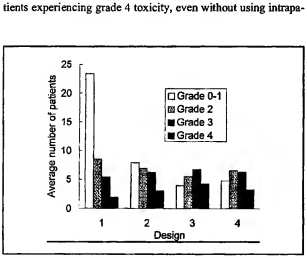

Fig. 4. Expected average number of patients in each toxicity grade for the 20 phase I trials studied. All patients are assumed to stay in the trial for three courses of therapy. Toxicity grade is the highest toxicity level experienced over the three courses.

The results presented above combined the conventional cohort escalation design 1 with the conventional intrapatient dose-modification option A. Combining design 1 with intrapatient option B has no effect on the number of patients or number of cohorts compared with 1A. It reduces the average number of patients with grade 0-1 as their highest level of toxicity from 23.3 to 19.3, but this is still not competitive with the numbers for designs 2 through 4 using option B.

Combining designs 2 through 4 with option A also has little or no effect on the number of patients or cohorts required compared with the same design using option B. In each case, about one fewer patient on average experiences grade 3 toxicity using option A than the same design with option B (5.2, 5.7, and 5.4 for 2A, 3A, and 4A, respectively). The expected number of patients with grade 4 toxicity is reduced on average by 0.4–1.1 patient (3.0, 4.3, and 3.2 for designs 2B, 3B, and 4B to 2.2, 3.2, and 2.8, respectively, for designs 2A, 3A, and 4A). The average number of patients with grade 0-1 toxicity is increased by about

2.4-3.0 patients on average (10.3, 6.5, and 7.0 for 2A, 3A, and 4A, respectively, compared with 7.9, 3.9, and 4.8). Much of the reduction in numbers of undertreated patients is achieved with designs 2A, 3A, and 4A compared with the standard design 1A, and they result in somewhat less grade 3-4 toxicity than the designs using dose titration. They are particularly attractive when there is preclinical concern about cumulative toxicity. They do not, however, provide patients accrued early in the trial a full opportunity to be treated at a dose that provides the greatest opportunity for benefit. Also, in situations where interpatient variability is substantial relative to $(K_2 - K_1)/\ln 1.4$ and intrapatient variability is small, designs without intrapatient dose escalation will not give each patient as much opportunity to be treated at a dose level appropriate to her particular level of drug tolerance and thus will be much less effective than designs with dose titration. Such combinations of parameters are not frequent in Table 1, but smaller values of σ_ε may be more prevalent with prospective use of accelerated designs.

Example

We generated one set of data for a clinical trial with the use of design 4 and the parameter values estimated from the actual data for trial T88-127 of chloroquinoxaline sulfonamide. Table 2 shows the data generated using these parameters. The first column lists patient sequence numbers. Each row of the table corresponds to a single patient. The numbers in a row represent the grades of toxicity experienced by that patient during her three courses of therapy. The columns correspond to dose levels, and the levels are labeled at the top of the columns.

The first patient received dose level 1 in her first course, and this resulted in toxicity grade 0 or 1. The table records this as a 0 because our simulations and analysis do not distinguish between grades 0 and 1. Since design 4 is used in this example, the first patient had her dose escalated by two steps for her second course, and she again showed grade 0-1 toxicity. Consequently, she received dose level 5 for her third course. She again showed no toxicity.

Since patient 1 had no toxicity in her first course, patient 2 started at dose level 3. Our simulations assumed that the time between patient entries was the same as the length of a single treatment course. Patient 2 also did not show any toxicity in her first course, and her dose was escalated two steps to level 5 for her second course. At that same time, patient 3 started at dose level 5.

The first toxicity observed was grade 2, which occurred in the second course of therapy for patient 4 at dose level 9. Hence, her dose was not escalated for her third course.

Patient 6 had grade 2 toxicity during her first course that was at dose level 11. She was kept at dose level 11 for her second course, but it resulted in no toxicity. Consequently, her dose was escalated to level 12 for her third course. It was escalated only a single dose step because the grade 2 toxicity she experienced during her first course was the second instance of grade 2 toxicity during the trial. This ended the rapid escalation phase of the design. Consequently, the cohort started at dose level 11 was expanded to three new patients started at that dose. The single dose escalations of three patients per cohort continued until the second patient started at dose level 15, patient 19, experienced grade 3 toxicity. That cohort is therefore expanded to six patients. Patient 22 experienced grade 4 toxicity in her first course at dose level 15, and hence the escalation of starting dose for new cohorts of patients stops. Three additional patients started

Table 2. Sequence of dose escalations and toxicity grades for patients treated in simulated phase I trial using design 4*

	Dose step†															
Patient No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0		0		0												
2			0		0		0										
3					0		0		0								
4							0_1		$2_2/0_3$								
5									0		0		0				
6											$2_1/0_2$	0_3					
7											0	0	0				
8											0	0	0				
9												0_1	0_2	2_3			
10												0_1	0_2	3_3			
11												0	0	0			
12													$2_1/2_2/2_3$				
13													0	0	0		
14													$2_1/2_2/2_3$				
15														0	0	0	
16														0_1	0_2	2_3	
17														0_1	$2_2/2_3$		
18															0	0_2	3_3
19														0_2	$3_1/2_3$		
20															$2_1/2_2/2_3$		
21															$2_1/2_2/0_3$		
22														$2_2/2_3$	4		
23															$2_1/2_2/0_3$		
24														0	0	0	
25														$0_1/2_3$	4_2		
26														0_1	0_2	2_3	

*Subscript = treatment course.

†Units are just the sequentially numbered dose steps. Level 1 = starting dose. Level 2 corresponds to a dose 40% greater than the starting dose.

on the next lower dose level, level 14. No patients experienced DLT at that level, and hence accrual to the trial was completed. The traditional recommended phase II dose would be level 14.

We fit the model to the data of Table 2 obtaining the following maximum likelihood estimates with 90% confidence intervals (CIs): K_1 is estimated as 7.4 (90% CI = 6.8-7.9) instead of the true value of 7.5; $(K_2 - K_1)/.34$ is estimated as 4.1 (90% CI = 1.3-7.0) instead of the true value of 4.6; $(K_3 - K_2)/0.34$ is estimated as 1.4 (90% CI = 0-2.9) instead of 2.9; α is estimated as 0 (90% CI = 0-0.67) with the true value of 0; σ_β is estimated as 0.71 (90% CI = 0.40-1.25), with a true value of 0.62; and σ_ε is estimated as 0.83 (90% CI = 0.37-1.84), with a true value of 0.90. In this example, there is good agreement between the estimates obtained from fitting the model and the true values used to generate the data example. The CIs are based on the usual normal approximations to the maximum likelihood estimates of the K's, $\log \sigma_\beta$, and $\log \sigma_\varepsilon$ and on the approximate chi-squared distribution of the logarithm of the likelihood ratio statistic as a function of α.

There is no evidence of cumulative toxicity because the alpha parameter is estimated as zero. There appears to be a substantial amount of both interpatient variability and intrapatient variability. The standard deviations are large, relative to the logarithm of the dose step ln(1.4) that is about 0.34. The K_1 and K_2 values appear to be well separated, but the K_2 and K_3 values are close.

Fig. 5 shows the probability of grade 2 or worse toxicity as a function of dose level and similar functions for the probability of grade 3 or worse toxicity and of grade 4 toxicity. These functions were computed by use of the model parameters estimated from the simulated data. From these graphs, one can estimate the dose level associated with any target level of any grade of toxicity. If one were to recommend a single dose level, the recommendation should reflect the distance between the grade 3+ curve and grade 4+ curve in Fig. 5. At dose level 17, the model estimates that 19% of the patients will experience grade 4 toxicity. At dose level 16, the probability of grade 4 toxicity is reduced to 12%, the probability of grade 3+ toxicity is 22%, and the probability of grade 2+ toxicity is 70%.

The functions in Fig. 5 do not give a clear picture of interpatient differences. Fig. 6 shows curves of the probability of grade 2+, 3+, and 4+ toxicity for three reprcsentative patients. The middle graph is for a patient whose β value equals the mean μ_β. The upper graph is for a patient whose β value is one standard deviation below the mean; i.e., $\mu_\beta - \sigma_\beta$. The lower graph is for a patient with $\beta = \mu_\beta + \sigma_\beta$. Dose levels 16 or 17 may be reasonable for the patient represented by the middle graph. For the patient represented by the upper graph, dose level 19 would be more appropriate. For the patient represented by the lower graph, dose level 14 or 15 would be more appropriate. This graph illustrates the substantial interpatient variability in the toxic response to this drug in this patient population. The separation between the grade 2+ and grade 3+ curves here and in Fig. 5 indicates the ability to effectively titrate patients to grade 2 toxicity. The closeness of the grade 3+ and grade 4+ curves indicates that doses that give grade 3 toxicity overlap substantially with those that give grade 4 toxicity. Use of any fixed dose for all patients is problematic, since any dose both overtreats and undertreats some patients. This is the principal conclusion of the data analysis.

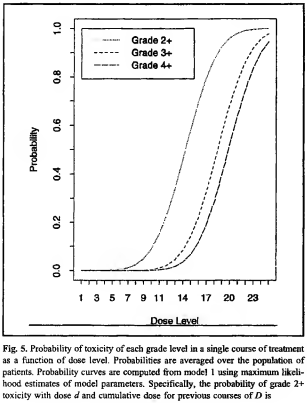

Fig. 5. Probability of toxicity of each grade level in a single course of treatment as a function of dose level. Probabilities are averaged over the population of patients. Probability curves are computed from model 1 using maximum likelihood estimates of model parameters. Specifically, the probability of grade 2+ toxicity with dose d and cumulative dose for previous courses of D is

$$\Phi\left(\frac{\log(d+\alpha D)+\mu_\beta - K_1}{\sqrt{\sigma_\beta^2+\sigma_\varepsilon^2}}\right),$$

where Φ denotes the cumulative standard normal distribution function. For computing probability of grade 3+ or grade 4 toxicity, replace K_1 by K_2 and K_3, respectively.

Discussion

The new designs described here appear to accomplish several objectives. They reduce the number of patients potentially undertreated. Some of these designs also reduce the duration of trials by doubling the dose until toxicity develops. These approaches also improve the information yield of phase 1 trials. They provide for estimation of the population distribution of the MTD and may also provide a statistical estimate of the degree of cumulative toxicity.

We have addressed phase 1 trials in which patients may receive more than one course of treatment. Not all phase 1 trials are of this type. Even in trials of this type, many patients remain in the study for only one or two courses of treatment because of tumor progression. This limits the information available for analysis. Patients may be able to remain in the study longer with accelerated titration designs because use of intrapatient dose escalation provides greater opportunity for therapeutic benefit. The reduced risk of design 4 compared with design 3 was based on using information from the second and third courses in determining when to stop the initial accelerated stage. This additional protection can be assured with fewer courses of treatment per patient by requiring that when the first instance of grade 2

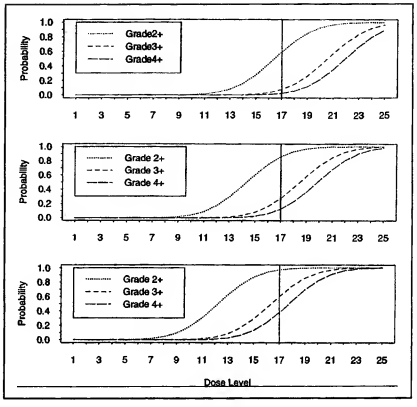

Fig. 6. Probability toxicity of each grade level in a single course of treatment as a function of dose level for individual patients. Probabilities are not averaged over the population of patients but are computed separately for the average patient (middle panel with $\beta_i = \mu_\beta$), the patient with reduced sensitivity to the toxic effects of the compound (upper panel with $\beta_i = \mu_\beta - \sigma_\beta$), and the patient with increased sensitivity to the toxic effects of the compound (lower panel with $\beta_i = \mu_\beta + \sigma_\beta$). Probability curves are computed from model 1 using maximum likelihood estimates of model parameters. Specifically, the probability of grade 2+ toxicity with dose d and cumulative dose for previous courses of D for a patient with value β_i is

$$\Phi\left(\frac{\log(d+\alpha D)+\beta_i-K_1}{\sigma_\varepsilon}\right),$$

where Φ denotes the cumulative standard normal distribution function. For computing probability of grade 3+ or grade 4 toxicity, replace K_1 by K_2 and K_3, respectively.

toxicity occurs, two other patients be treated at that same dose without grade 2 toxicity before the dose is doubled. This may be satisfied by later courses at escalated doses in previous patients or may require starting a new patient at the same dose as the one who experienced grade 2 toxicity. This modification is not needed for design 2, since it uses only first-course toxicity for determining when to terminate the accelerated stage and uses smaller dose steps. For designs without intrapatient dose escalation, this modification would increase the number of patients treated at lower doses and may extend the time to completion.

In these simulations, we used the conventional stopping rule with all designs for consistency. The study stopped when two patients experienced DLT at a dose level, and six patients were treated at the next lower dose level with no more than one patient experiencing DLT. For the new designs, the population distribution of MTDs is estimated, and there is nothing special about the highest dose at which fewer than two patients experienced DLT. One might, therefore, continue entering patients beyond the usual stopping point to refine the estimates of the population distributions. In fact, the entire second stage of sampling could use a model-based or Bayesian approach to selecting the first-course dose for each patient. Simple up-down phase I designs with cohort sizes other than three to six patients are also sometimes used when the amount of DLT that is to be tolerated is much less than 33%.

We have analyzed the results of 20 actual phase I trials by use of the model described above for the generation of simulated trial data. Other models could be used in place of the expression shown in equation 1 and, in particular, other approaches to the modeling of cumulative toxicity may be more appropriate in specific trials.

Use of an accelerated titration design requires careful definition of the level of toxicity considered dose limiting and the level considered sufficiently low (e.g., none-mild) that intrapatient dose escalation is acceptable. These definitions must be made for each organ system. In the simulations, we tabulated the incidence of unacceptable or grade 4 toxicity, but this is not necessary in using an accelerated titration design. The dose escalation and de-escalation decisions that must be made during the trial depend on distinguishing none-mild toxicity from moderate toxicity and on distinguishing moderate toxicity from DLT. These definitions may be protocol specific. The tracking of toxicity over multiple treatment courses and the use of intrapatient titrations require careful patient management. However, the result will enhance the likelihood that patients receive therapeutic dosing and increase the useful information obtained from each treated patient.

The approach to design and analysis of phase I trials described in this article will help identify when there is large interpatient variability in sensitivity to the toxic effects of a drug. If interpatient variability is small, a fixed-dose regimen can be used in phase II trials, and few patients will be either overdosed or underdosed. Mick et al. (*18*) have described important sources of interpatient and intrapatient variability that might be usefully incorporated into the model. Further improvement might result from modeling toxicity separately by organ system.

Pharmacokinetic differences are sometimes an important source of interpatient variability. In such cases, it may be advis-

able to attempt to control systemic exposure rather than dose. If drug clearance can be predicted by use of baseline patient characteristics such as liver or renal function, then the dose needed to achieve the targeted concentration can be estimated. Otherwise, an adaptive dosing scheme may be needed to achieve a target exposure. If the drug is delivered by a prolonged infusion, one may adapt the infusion rate based on estimates of pharmacokinetic parameters to target systemic exposure levels. The accelerated titration designs described here then may be applied with the only change being the use of exposure levels rather than dose levels. When prolonged infusions are not used, it may not be feasible to deliver a target exposure level during the same course of treatment in which pharmacokinetic parameters are estimated. It still may be possible, however, to use parameters estimated in the first course of treatment for the titration of exposure in subsequent courses.

Accelerated titration designs are more aggressive than standard approaches and, therefore, may be associated with more risk. The simulations were performed with a very wide range of model parameters and suggest that the risks appear acceptable for designs 2 and 4. We believe that these designs are appropriate for clinical testing. For drugs that exhibit preclinical evidence of cumulative toxicity, special caution in the conduct of any type of phase I trial is needed. Accelerated designs without intrapatient dose escalations achieve most of the advantages of accelerated titration designs, with little or no increase in risk compared with the standard design 1A. However, they do not provide as great a reduction in the number of undertreated patients and, in particular, do not provide patients accrued early in the trial or those who have an especially high individual tolerance for the drug as much opportunity as do titration designs to be treated at a dose that provides the greatest opportunity for benefit. We hope to sponsor phase I clinical trials to provide prospective evaluation of these new approaches.

Appendix

Four designs were evaluated as follows. **Design 1:** cohorts of three new patients per dose level. If one of three patients experiences DLT in the first course, expand the cohort to six patients. Intrapatient escalation option A. **Design 2:** cohorts of one new patient per dose level. When the first instance of first course DLT is observed, or the second instance of first course grade 2 toxicity of any type, expand the cohort for current dose level and revert to use of design 1 for all further cohorts. Intrapatient escalation option B. **Design 3:** same as design 2, except that double dose steps are used during the initial accelerated stage of the trial (both for between-patient and within-patient escalations). Intrapatient escalation option B. **Design 4:** cohorts of one new patient per dose level and double dose steps are used during the initial accelerated stage of the trial. When the first instance of DLT is observed at any course or the second instance of any course grade 2 toxicity of any type, expand the cohort for current dose level and revert to use of design 1 for all further cohorts. Intrapatient escalation option B.

The intrapatient dose modification options are defined as follows:

Option A: no within-patient dose escalation. De-escalate if grade 3 or worse toxicity at previous course.

Option B: Escalate if grade 0-1 toxicity at previous course. De-escalate if grade 3 or worse toxicity at previous course.

References

(1) Ratain MJ, Mick R, Schilsky RL, Siegler M. Statistical and ethical issues in the design and conduct of phase I and II clinical trials of new anticancer agents. J Natl Cancer Inst 1993;85:1637-43.

(2) Estey E, Hoth D, Simon R, Marsoni S, Leyland-Jones B, Wittes R. Therapeutic response in phase I trials of antineoplastic agents. Cancer Treat Rep 1986;70:1105-15.

(3) Collins JM, Zaharko DS, Dedrick RL, Chabner BA. Potential roles for preclinical pharmacology in phase I clinical trials. Cancer Treat Rep 1986; 70:73-80.

(4) Storer BE. Design and analysis of phase I clinical trials. Biometrics 1989; 45:925-37.

(5) Eichhorn BH, Zacks S. Sequential search for an optimal dosage. J Am Stat Assoc 1973;68:594-8.

(6) O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. Biometrics 1990;46:33-48.

(7) Faries D. The modified continual reassessment method for phase I cancer clinical trials. American Statistical Association 1991 Proceedings of the Biopharmaceutical Section. Alexandria (VA): American Statistical Association, 1991:269-73.

(8) Goodman SN, Zahurak ML, Piantadosi S. Some practical improvements in the continual reassessment method for phase I studies. Stat Med 1995;14: 1149-61.

(9) Moller S. An extension of the continual reassessment methods using a preliminary up-and-down design in a dose finding study in cancer patients, in order to investigate a greater range of doses. Stat Med 1995;14: 911-22.

(10) Piantadosi S, Liu G. Improved designs for dose escalation studies using pharmacokinetic measurements. Stat Med 1996;15:1605-18.

(11) Mick R, Ratain MJ. Model-guided determination of maximum tolerated dose in phase 1 clinical trials: evidence for increased precision. J Natl Cancer Inst 1993;85:217-23.

(12) Sheiner LB, Beal SL, Sambol NC. Study designs for dose-ranging. Clin Pharmacol Ther 1989;46:63-77.

(13) Sheiner LB, Hashimoto Y, Beal SL. A simulation study comparing designs for dose ranging. Stat Med 1991;10:303-21.

(14) Chuang C. The analysis of a titration study. Stat Med 1987;6:583-90.

(15) Shih WJ, Gould AL, Hwang IK. The analysis of titration studies in phase III clinical trials. Stat Med 1989;8:583-91.

(16) Chou TC, Talalay P. Generalized equations for the analysis of inhibitions of Michaelis-Menten and higher-order kinetic systems with two or more mutually exclusive and nonexclusive inhibitors. Eur J Biochem 1981;115: 207-16.

(17) Benet L, Oie S, Schwartz J. Design and optimization of dosage regimens; pharmacokinetic data. In: Hardman J, Limbird L, Molinoff P, et al., editors. The pharmacological basis of therapeutics. New York: McGraw-Hill, 1996: 1707-92.

(18) Mick R, Lane N, Daugherty C, Ratain MJ. Physician-determined patient risk of toxic effects: impact on enrollment and decision making in phase I cancer trials. J Natl Cancer Inst 1994;86:1685-93.

Notes

We thank Michelle Gossard and Sally Lopes of The Emmes Corporation for their assistance in the management of the phase I trial database used for this study. We also thank the reviewers for suggestions on improvement of an earlier draft of the manuscript.

Manuscript received January 27, 1997; revised May 16, 1997; accepted May 30, 1997.

Attach 4 CR 264

SPECIAL ARTICLE

New Guidelines to Evaluate the Response to Treatment in Solid Tumors

Patrick Therasse, Susan G. Arbuck, Elizabeth A. Eisenhauer, Jantien Wanders, Richard S. Kaplan, Larry Rubinstein, Jaap Verweij, Martine Van Glabbeke, Allan T. van Oosterom, Michaele C. Christian, Steve G. Gwyther

Anticancer cytotoxic agents go through a process by which their antitumor activity—on the basis of the amount of tumor shrinkage they could generate—has been investigated. In the late 1970s, the International Union Against Cancer and the World Health Organization introduced specific criteria for the codification of tumor response evaluation. In 1994, several organizations involved in clinical research combined forces to tackle the review of these criteria on the basis of the experience and knowledge acquired since then. After several years of intensive discussions, a new set of guidelines is ready that will supersede the former criteria. In parallel to this initiative, one of the participating groups developed a model by which response rates could be derived from unidimensional measurement of tumor lesions instead of the usual bidimensional approach. This new concept has been largely validated by the Response Evaluation Criteria in Solid Tumors Group and integrated into the present guidelines. This special article also provides some philosophic background to clarify the various purposes of response evaluation. It proposes a model by which a combined assessment of all existing lesions, characterized by target lesions (to be measured) and nontarget lesions, is used to extrapolate an overall response to treatment. Methods of assessing tumor lesions are better codified, briefly within the guidelines and in more detail in Appendix I. All other aspects of response evaluation have been discussed, reviewed, and amended whenever appropriate. [J Natl Cancer Inst 2000; 92:205–16]

A. Preamble

Early attempts to define the objective response of a tumor to an anticancer agent were made in the early 1960s *(1,2)*. In the mid- to late 1970s, the definitions of objective tumor response were widely disseminated and adopted when it became apparent that a common language would be necessary to report the results of cancer treatment in a consistent manner.

The World Health Organization (WHO) definitions published in the 1979 *WHO Handbook (3)* and by Miller et al. *(4)* in 1981 have been the criteria most commonly used by investigators around the globe. However, some problems have developed with the use of WHO criteria: 1) The methods for integrating into response assessments the change in size of measurable and "evaluable" lesions as defined by WHO vary among research groups, 2) the minimum lesion size and number of lesions to be recorded also vary, 3) the definitions of progressive disease are related to change in a single lesion by some and to a change in the overall tumor load (sum of the measurements of all lesions) by others, and 4) the arrival of new technologies (computed tomography [CT] and magnetic resonance imaging [MRI]) has led to some confusion about how to integrate three-dimensional measures into response assessment.

These issues and others have led to a number of different modifications or clarifications to the WHO criteria, resulting in a situation where response criteria are no longer comparable among research organizations—the very circumstance that the WHO publication had set out to avoid. This situation led to an initiative undertaken by representatives of several research groups to review the response definitions in use and to create a revision of the WHO criteria that, as far as possible, addressed areas of conflict and inconsistency.

In so doing, a number of principles were identified:

1) Despite the fact that "novel" therapies are being developed that may work by mechanisms unlikely to cause tumor regression, there remains an important need to continue to describe objective change in tumor size in solid tumors for the foreseeable future. Thus, the four categories of complete response, partial response, stable disease, and progressive disease, as originally categorized in the *WHO Handbook (3)*, should be retained in any new revision.
2) Because of the need to retain some ability to compare favorable results of future therapies with those currently available, it was agreed that no major discrepancy in the meaning and the concept of partial response should exist between the old and the new guidelines, although measurement criteria would be different.
3) In some institutions, the technology now exists to determine

Affiliations of authors: P. Therasse, J. Verweij, M. Van Glabbeke, A. T. van Oosterom, European Organization for Research and Treatment of Cancer, Brussels, Belgium; S. G. Arbuck, R. S. Kaplan, L. Rubinstein, M. C. Christian, National Cancer Institute, Bethesda, MD; E. A. Eisenhauer, National Cancer Institute of Canada Clinical Trials Group, Kingston, ON, Canada; J. Wanders, New Drug Development Office Oncology, Amsterdam, The Netherlands; S. G. Gwyther, East Surrey Healthcare National Health Service Trust, Redhill, U.K.

Correspondence to: Patrick Therasse, M.D., European Organization for Research and Treatment of Cancer Data Center, Avenue Mounier 83/11, 1200 Brussels, Belgium (e-mail: pth@eortc.be).

See "Note" following "References."

changes in tumor volume or changes in tumor metabolism that may herald shrinkage. However, these techniques are not yet widely available, and many have not been validated. Furthermore, it was recognized that the utility of response criteria to date had not been related to precision of measurement. The definition of a partial response, in particular, is an arbitrary convention—there is no inherent meaning for an individual patient of a 50% decrease in overall tumor load. It was not thought that increased precision of measurement of tumor volume was an important goal for its own sake. Rather, standardization and simplification of methodology were desirable. Nevertheless, the guidelines proposed in this document are not meant to discourage the development of new tools that may provide more reliable surrogate end points than objective tumor response for predicting a potential therapeutic benefit for cancer patients.

4) Concerns regarding the case with which a patient may be considered mistakenly to have disease progression by the current WHO criteria (primarily because of measurement error) have already led some groups such as the Southwest Oncology Group to adopt criteria that require a greater increase in size of the tumor to consider a patient to have progressive disease *(5)*. These concerns have led to a similar change within these revised WHO criteria (*see* Appendix II).

5) These criteria have not addressed several other areas of recent concern, but it is anticipated that this process will continue and the following will be considered in the future:
   - Measures of antitumor activity, other than tumor shrinkage, that may appropriately allow investigation of cytostatic agents in phase II trials;
   - Definitions of serum marker response and recommended methodology for their validation; and
   - Specific tumors or anatomic sites presenting unique complexities.

B. Background

These guidelines are the result of a large, international collaboration. In 1994, the European Organization for Research and Treatment of Cancer (EORTC), the National Cancer Institute (NCI) of the United States, and the National Cancer Institute of Canada Clinical Trials Group set up a task force (*see* Appendix III) with the main objective of reviewing the existing sets of criteria used to evaluate response to treatment in solid tumors. A 3 years of regular meetings and exchange of ideas within the task force, a draft revised version of the WHO criteria was produced and widely circulated (*see* Appendix IV). Comments received (response rate, 95%) were compiled and discussed within the task force before a second version of the document integrating relevant comments was issued. This second version of the document was again circulated to external reviewers who were also invited to participate in a consensus meeting (on behalf of the organization that they represented) to discuss and finalize unresolved problems (October 1998). The list of participants to this consensus meeting is shown in Appendix IV and included representatives from academia, industry, and regulatory authorities. Following the recommendations discussed during the consensus meeting, a third version of the document was produced, presented publicly to the scientific community (American Society for Clinical Oncology, 1999), and submitted to the *Journal of the National Cancer Institute* in June 1999 for official publication.

Data from collaborative studies, including more than 4000 patients assessed for tumor response, support the simplification of response evaluation through the use of unidimensional measurements and the sum of the longest diameters instead of the conventional method using two measurements and the sum of the products. The results of the different retrospective analyses (comparing both approaches) performed by use of these different databases are described in Appendix V. This new approach, which has been implemented in the following guidelines, is based on the model proposed by James et al. *(6)*.

C. Response Evaluation Criteria in Solid Tumors (RECIST) Guidelines

1. Introduction

The introduction explores the definitions, assumptions, and purposes of tumor response criteria. Below, guidelines that are offered may lead to more uniform reporting of outcomes of clinical trials. Note that, although single investigational agents are discussed, the principles are the same for drug combinations, noninvestigational agents, or approaches that do not involve drugs.

Tumor response associated with the administration of anticancer agents can be evaluated for at least three important purposes that are conceptually distinct:

- Tumor response as a prospective end point in early clinical trials. In this situation, objective tumor response is employed to determine whether the agent/regimen demonstrates sufficiently encouraging results to warrant further testing. These trials are typically phase II trials of investigational agents/regimens (*see* section 1.2), and it is for use in this precise context that these guidelines have been developed.
- Tumor response as a prospective end point in more definitive clinical trials designed to provide an estimate of benefit for a specific cohort of patients. These trials are often randomized comparative trials or single-arm comparisons of combinations of agents with historical control subjects. In this setting, objective tumor response is used as a surrogate end point for other measures of clinical benefit, including time to event (death or disease progression) and symptom control (*see* section 1.3).
- Tumor response as a guide for the clinician and patient or study subject in decisions about continuation of current therapy. This purpose is applicable both to clinical trials and to routine practice (*see* section 1.1), but use in the context of decisions regarding continuation of therapy is not the primary focus of this document.

However, in day-to-day usage, the distinction among these uses of the term "tumor response" can easily be missed, unless an effort is made to be explicit. When these differences are ignored, inappropriate methodology may be used and incorrect conclusions may result.

1.1. Response Outcomes in Daily Clinical Practice of Oncology

The evaluation of tumor response in the daily clinical practice of oncology may not be performed according to predefined criteria. It may, rather, be based on a subjective medical judgment that results from clinical and laboratory data that are used to assess the treatment benefit for the patient. The defined criteria

developed further in this document are not necessarily applicable or complete in such a context. It might be appropriate to make a distinction between "clinical improvement" and "objective tumor response" in routine patient management outside the context of a clinical trial.

1.2. Response Outcomes in Uncontrolled Trials as a Guide to Further Testing of a New Therapy

"Observed response rate" is often employed in single-arm studies as a "screen" for new anticancer agents that warrant further testing. Related outcomes, such as response duration or proportion of patients with complete responses, are sometimes employed in a similar fashion. The utilization of a response rate in this way is not encumbered by an implied assumption about the therapeutic benefit of such responses but rather implies some degree of biologic antitumor activity of the investigated agent.

For certain types of agents (i.e., cytotoxic drugs and hormones), experience has demonstrated that objective antitumor responses observed at a rate higher than would have been expected to occur spontaneously can be useful in selecting anticancer agents for further study. Some agents selected in this way have eventually proven to be clinically useful. Furthermore, criteria for "screening" new agents in this way can be modified by accumulated experience and eventually validated in terms of the efficiency by which agents so screened are shown to be of clinical value by later, more definitive, trials.

In most circumstances, however, a new agent achieving a response rate determined *a priori* to be sufficiently interesting to warrant further testing may not prove to be an effective treatment for the studied disease in subsequent randomized phase III trials. Random variables and selection biases, both known and unknown, can have an overwhelming effect in small, uncontrolled trials. These trials are an efficient and economic step for initial evaluation of the activity of a new agent or combination in a given disease setting. However, many such trials are performed, and the proportion that will provide false-positive results is necessarily substantial. In many circumstances, it would be appropriate to perform a second small confirmatory trial before initiating large resource-intensive phase III trials.

Sometimes, several new therapeutic approaches are studied in a randomized phase II trial. The purpose of randomization in this setting, as in phase III studies, is to minimize the impact of random imbalances in prognostic variables. However, randomized phase II studies are, by definition, not intended to provide an adequately powered comparison between arms (regimens). Rather, the goal is simply to identify one or more arms for further testing, and the sample size is chosen so to provide reasonable confidence that a truly inferior arm is not likely to be selected. Therefore, reporting the results of such randomized phase II trials should not imply statistical comparisons between treatment arms.

1.3. Response Outcomes in Clinical Trials as a Surrogate for Palliative Effect

1.3.1. Use in nonrandomized clinical trials. The only circumstance in which objective responses in a nonrandomized trial can permit a tentative assumption of a palliative effect (i.e., beyond a purely clinical measure of benefit) is when there is an actual or implied comparison with historical series of similar patients. This assumption is strongest when the prospectively determined statistical analysis plan provides for matching of relevant prognostic variables between case subjects and a defined series of control subjects. Otherwise, there must be, at the very least, prospectively determined statistical criteria that provide a very strong justification for assumptions about the response rate that would have been expected in the appropriate "control" population (untreated or treated with conventional therapy, as fits the clinical setting). However, even under these circumstances, a high rate of observed objective response does not constitute proof or confirmation of clinical therapeutic benefit. Because of unavoidable and nonquantifiable biases inherent in nonrandomized trials, proof of benefit still requires eventual confirmation in a prospectively randomized, controlled trial of adequate size. The appropriate end points of therapeutic benefit for such a trial are survival, progression-free survival, or symptom control (including quality of life).

1.3.2. Use in randomized trials. Even in the context of prospectively randomized phase III comparative trials, "observed response rate" should not be the sole, or major, end point. The trial should be large enough that differences in response rate can be validated by association with more definitive end points reflecting therapeutic benefit, such as survival, progression-free survival, reduction in symptoms, or improvement (or maintenance) of quality of life.

2. Measurability of Tumor Lesions at Baseline

2.1. Definitions

At baseline, tumor lesions will be categorized as follows: measurable (lesions that can be accurately measured in at least one dimension [longest diameter to be recorded] as ≥20 mm with conventional techniques or as ≥10 mm with spiral CT scan [*see* section 2.2]) or nonmeasurable (all other lesions, including small lesions [longest diameter <20 mm with conventional techniques or <10 mm with spiral CT scan] and truly nonmeasurable lesions).

The term "evaluable" in reference to measurability is not recommended and will not be used because it does not provide additional meaning or accuracy.

All measurements should be recorded in metric notation by use of a ruler or calipers. All baseline evaluations should be performed as closely as possible to the beginning of treatment and never more than 4 weeks before the beginning of treatment.

Lesions considered to be truly nonmeasurable include the following: bone lesions, leptomeningeal disease, ascites, pleural/pericardial effusion, inflammatory breast disease, lymphangitis cutis/pulmonis, abdominal masses that are not confirmed and followed by imaging techniques, and cystic lesions.

(*Note:* Tumor lesions that are situated in a previously irradiated area might or might not be considered measurable, and the conditions under which such lesions should be considered must be defined in the protocol when appropriate.)

2.2. Specifications by Methods of Measurements

The same method of assessment and the same technique should be used to characterize each identified and reported lesion at baseline and during follow-up. Imaging-based evaluation is preferred to evaluation by clinical examination when both methods have been used to assess the antitumor effect of a treatment.

2.2.1. Clinical examination. Clinically detected lesions will only be considered measurable when they are superficial (e.g., skin nodules and palpable lymph nodes). For the case of skin lesions, documentation by color photography—including a ruler to estimate the size of the lesion—is recommended.

2.2.2. Chest x-ray. Lesions on chest x-ray are acceptable as measurable lesions when they are clearly defined and surrounded by aerated lung. However, CT is preferable. More details concerning the use of this method of assessment for objective tumor response evaluation are provided in Appendix I.

2.2.3. CT and MRI. CT and MRI are the best currently available and most reproducible methods for measuring target lesions selected for response assessment. Conventional CT and MRI should be performed with contiguous cuts of 10 mm or less in slice thickness. Spiral CT should be performed by use of a 5-mm contiguous reconstruction algorithm; this specification applies to the tumors of the chest, abdomen, and pelvis, while head and neck tumors and those of the extremities usually require specific protocols. More details concerning the use of these methods of assessment for objective tumor response evaluation are provided in Appendix I.

2.2.4. Ultrasound. When the primary end point of the study is objective response evaluation, ultrasound should not be used to measure tumor lesions that are clinically not easily accessible. It may be used as a possible alternative to clinical measurements for superficial palpable lymph nodes, subcutaneous lesions, and thyroid nodules. Ultrasound might also be useful to confirm the complete disappearance of superficial lesions usually assessed by clinical examination. Justifications for not using ultrasound to measure tumor lesions for objective response evaluation are provided in Appendix I.

2.2.5. Endoscopy and laparoscopy. The utilization of these techniques for objective tumor evaluation has not yet been fully or widely validated. Their uses in this specific context require sophisticated equipment and a high level of expertise that may be available only in some centers. Therefore, utilization of such techniques for objective tumor response should be restricted to validation purposes in specialized centers. However, such techniques can be useful in confirming complete histopathologic response when biopsy specimens are obtained.

2.2.6. Tumor markers. Tumor markers alone cannot be used to assess response. However, if markers are initially above the upper normal limit, they must return to normal levels for a patient to be considered in complete clinical response when all tumor lesions have disappeared. Specific additional criteria for standardized usage of prostate-specific antigen and CA (cancer antigen) 125 response in support of clinical trials are being validated.

2.2.7. Cytology and histology. Cytologic and histologic techniques can be used to differentiate between partial response and complete response in rare cases (e.g., after treatment to differentiate between residual benign lesions and residual malignant lesions in tumor types such as germ cell tumors). Cytologic confirmation of the neoplastic nature of any effusion that appears or worsens during treatment is required when the measurable tumor has met criteria for response or stable disease. Under such circumstances, the cytologic examination of the fluid collected will permit differentiation between response or stable disease (an effusion may be a side effect of the treatment) and progressive disease (if the neoplastic origin of the fluid is confirmed). New techniques to better establish objective tumor response will be integrated into these criteria when they are fully validated to be used in the context of tumor response evaluation.

3. Tumor Response Evaluation

3.1. Baseline Evaluation

3.1.1. Assessment of overall tumor burden and measurable disease. To assess objective response, it is necessary to estimate the overall tumor burden at baseline to which subsequent measurements will be compared. Only patients with measurable disease at baseline should be included in protocols where objective tumor response is the primary end point. Measurable disease is defined by the presence of at least one measurable lesion (as defined in section 2.1). If the measurable disease is restricted to a solitary lesion, its neoplastic nature should be confirmed by cytology/histology.

3.1.2. Baseline documentation of "target" and "nontarget" lesions. All measurable lesions up to a maximum of five lesions per organ and 10 lesions in total, representative of all involved organs, should be identified as target lesions and recorded and measured at baseline. Target lesions should be selected on the basis of their size (those with the longest diameter) and their suitability for accurate repeated measurements (either by imaging techniques or clinically). A sum of the longest diameter for all target lesions will be calculated and reported as the baseline sum longest diameter. The baseline sum longest diameter will be used as the reference by which to characterize the objective tumor response.

All other lesions (or sites of disease) should be identified as nontarget lesions and should also be recorded at baseline. Measurements of these lesions are not required, but the presence or absence of each should be noted throughout follow-up.

3.2. Response Criteria

3.2.1. Evaluation of target lesions. This section provides the definitions of the criteria used to determine objective tumor response for target lesions. The criteria have been adapted from the original *WHO Handbook (3),* taking into account the measurement of the longest diameter only for all target lesions: complete response—the disappearance of all target lesions; partial response—at least a 30% decrease in the sum of the longest diameter of target lesions, taking as reference the baseline sum longest diameter; progressive disease—at least a 20% increase in the sum of the longest diameter of target lesions, taking as reference the smallest sum longest diameter recorded since the treatment started or the appearance of one or more new lesions; stable disease—neither sufficient shrinkage to qualify for partial response nor sufficient increase to qualify for progressive disease, taking as reference the smallest sum longest diameter since the treatment started.

3.2.2. Evaluation of nontarget lesions. This section provides the definitions of the criteria used to determine the objective tumor response for nontarget lesions: complete response—the disappearance of all nontarget lesions and normalization of tumor marker level; incomplete response/stable disease—the persistence of one or more nontarget lesion(s) and/or the maintenance of tumor marker level above the normal limits; and progressive disease—the appearance of one or more new lesions and/or unequivocal progression of existing nontarget lesions *(1).*

(Note: Although a clear progression of "nontarget" lesions only is exceptional, in such circumstances, the opinion of the

treating physician should prevail and the progression status should be confirmed later by the review panel [or study chair]).

3.2.3. Evaluation of best overall response. The best overall response is the best response recorded from the start of treatment until disease progression/recurrence (taking as reference for progressive disease the smallest measurements recorded since the treatment started). In general, the patient's best response assignment will depend on the achievement of both measurement and confirmation criteria (*see* section 3.3.1). Table 1 provides overall responses for all possible combinations of tumor responses in target and nontarget lesions with or without the appearance of new lesions.

(*Notes:*

- Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be classified as having "symptomatic deterioration." Every effort should be made to document the objective disease progression, even after discontinuation of treatment.
- Conditions that may define early progression, early death, and inevaluability are study specific and should be clearly defined in each protocol (depending on treatment duration and treatment periodicity).
- In some circumstances, it may be difficult to distinguish residual disease from normal tissue. When the evaluation of complete response depends on this determination, it is recommended that the residual lesion be investigated (fine-needle aspiration/biopsy) before confirming the complete response status.)

3.2.4. Frequency of tumor re-evaluation. Frequency of tumor re-evaluation while on treatment should be protocol specific and adapted to the type and schedule of treatment. However, in the context of phase II studies where the beneficial effect of therapy is not known, follow-up of every other cycle (i.e., 6–8 weeks) seems a reasonable norm. Smaller or greater time intervals than these could be justified in specific regimens or circumstances.

After the end of the treatment, the need for repetitive tumor evaluations depends on whether the phase II trial has, as a goal, the response rate or the time to an event (disease progression/death). If time to an event is the main end point of the study, then routine re-evaluation is warranted of those patients who went off therapy for reasons other than the expected event at frequencies to be determined by the protocol. Intervals between evaluations twice as long as on study are often used, but no strict rule can be made.

Table 1. Overall responses for all possible combinations of tumor responses in target and nontarget lesions with or without the appearance of new lesions*

Target lesions	Nontarget lesions	New lesions	Overall response
CR	CR	No	CR
CR	Incomplete response/SD	No	PR
PR	Non-PD	No	PR
SD	Non-PD	No	SD
PD	Any	Yes or no	PD
Any	PD	Yes or no	PD
Any	Any	Yes	PD

*CR = complete response; PR = partial response; SD = stable disease; and PD = progressive disease. *See* text for more details.

3.3. Confirmatory Measurement/Duration of Response

3.3.1. Confirmation. The main goal of confirmation of objective response in clinical trials is to avoid overestimating the response rate observed. This aspect of response evaluation is particularly important in nonrandomized trials where response is the primary end point. In this setting, to be assigned a status of partial response or complete response, changes in tumor measurements must be confirmed by repeat assessments that should be performed no less than 4 weeks after the criteria for response are first met. Longer intervals as determined by the study protocol may also be appropriate.

In the case of stable disease, measurements must have met the stable disease criteria at least once after study entry at a minimum interval (in general, not less than 6–8 weeks) that is defined in the study protocol (*see* section 3.3.3).

(*Note:* Repeat studies to confirm changes in tumor size may not always be feasible or may not be part of the standard practice in protocols where progression-free survival and overall survival are the key end points. In such cases, patients will not have "confirmed response." This distinction should be made clear when reporting the outcome of such studies.)

3.3.2. Duration of overall response. The duration of overall response is measured from the time that measurement criteria are met for complete response or partial response (whichever status is recorded first) until the first date that recurrent or progressive disease is objectively documented (taking as reference for progressive disease the smallest measurements recorded since the treatment started). The duration of overall complete response is measured from the time measurement criteria are first met for complete response until the first date that recurrent disease is objectively documented.

3.3.3. Duration of stable disease. Stable disease is measured from the start of the treatment until the criteria for disease progression is met (taking as reference the smallest measurements recorded since the treatment started). The clinical relevance of the duration of stable disease varies for different tumor types and grades. Therefore, it is highly recommended that the protocol specify the minimal time interval required between two measurements for determination of stable disease. This time interval should take into account the expected clinical benefit that such a status may bring to the population under study.

(*Note:* The duration of response or stable disease as well as the progression-free survival are influenced by the frequency of follow-up after baseline evaluation. It is not in the scope of this guideline to define a standard follow-up frequency that should take into account many parameters, including disease types and stages, treatment periodicity, and standard practice. However, these limitations to the precision of the measured end point should be taken into account if comparisons among trials are to be made.)

3.4. Progression-Free Survival/Time to Progression

This document focuses primarily on the use of objective response end points. In some circumstances (e.g., brain tumors or investigation of noncytoreductive anticancer agents), response evaluation may not be the optimal method to assess the potential anticancer activity of new agents/regimens. In such cases, progression-free survival/time to progression can be considered valuable alternatives to provide an initial estimate of biologic effect of new agents that may work by a noncytotoxic mecha-

nism. It is clear though that, in an uncontrolled trial proposing to utilize progession-free survival/time to progression, it will be necessary to document with care the basis for estimating what magnitude of progression-free survival/time to progression would be expected in the absence of a treatment effect. It is also recommended that the analysis be quite conservative in recognition of the likelihood of confounding biases, e.g., with regard to selection and ascertainment. Uncontrolled trials using progression-free survival or time to progression as a primary end point should be considered on a case-by-case basis, and the methodology to be applied should be thoroughly described in the protocol.

4. Response Review

For trials where the response rate is the primary end point, it is strongly recommended that all responses be reviewed by an expert or experts independent of the study at the study's completion. Simultaneous review of the patients' files and radiologic images is the best approach.

(*Note:* When a review of the radiologic images is to take place, it is also recommended that images be free of marks that might obscure the lesions or bias the evaluation of the reviewer[s]).

5. Reporting of Results

All patients included in the study must be assessed for response to treatment, even if there are major protocol treatment deviations or if they are ineligible. Each patient will be assigned one of the following categories: 1) complete response, 2) partial response, 3) stable disease, 4) progressive disease, 5) early death from malignant disease, 6) early death from toxicity, 7) early death because of other cause, or 9) unknown (not assessable, insufficient data). (*Note:* By arbitrary convention, category 9 usually designates the "unknown" status of any type of data in a clinical database.)

All of the patients who met the eligibility criteria should be included in the main analysis of the response rate. Patients in response categories 4–9 should be considered as failing to respond to treatment (disease progression). Thus, an incorrect treatment schedule or drug administration does not result in exclusion from the analysis of the response rate. Precise definitions for categories 4–9 will be protocol specific.

All conclusions should be based on all eligible patients.

Subanalyses may then be performed on the basis of a subset of patients, excluding those for whom major protocol deviations have been identified (e.g., early death due to other reasons, early discontinuation of treatment, major protocol violations, etc). However, these subanalyses may not serve as the basis for drawing conclusions concerning treatment efficacy, and the reasons for excluding patients from the analysis should be clearly reported. The 95% confidence intervals should be provided.

6. Response Evaluation in Randomized Phase III Trials

Response evaluation in phase III trials may be an indicator of the relative antitumor activity of the treatments evaluated but may usually not solely predict the real therapeutic benefit for the population studied. If objective response is selected as a primary end point for a phase III study (only in circumstances where a direct relationship between objective tumor response and a real therapeutic benefit can be unambiguously demonstrated for the population studied), the same criteria as those applicable to phase II trials (RECIST guidelines) should be used.

On the other hand, some of the guidelines presented in this special article might not be required in trials, such as phase III trials, in which objective response is *not* the primary end point. For example, in such trials, it might not be necessary to measure as many as 10 target lesions or to confirm response with a follow-up assessment after 4 weeks or more. Protocols should be written clearly with respect to planned response evaluation and whether confirmation is required so as to avoid *post-hoc* decisions affecting patient evaluability.

Appendix I. Specifications for Radiologic Imaging

These notes are recommendations for use in clinical studies and, as such, these protocols for computed tomography (CT) and magnetic resonance imaging (MRI) scanning may differ from those employed in clinical practice at various institutions. The use of standardized protocols allows comparability both within and between different studies, irrespective of where the examination has been undertaken.

Specific Notes

- For chest x-ray, not only should the film be performed in full inspiration in the posteroanterior projection, but also the film to tube distance should remain constant between examinations. However, patients in trials with advanced disease may not be well enough to fulfill these criteria, and such situations should be reported together with the measurements.

Lesions bordering the thoracic wall are not suitable for measurements by chest x-ray, since a slight change in position of the patients can cause considerable differences in the plane in which the lesion is projected and may appear to cause a change that is actually an artifact. These lesions should be followed by a CT or an MRI. Similarly, lesions bordering or involving the mediastinum should be documented on CT or MRI.

- CT scans of the thorax, abdomen, and pelvis should be contiguous throughout the anatomic region of interest. As a rule of thumb, the minimum size of the lesion should be no less than double the slice thickness. Lesions smaller than this are subject to substantial "partial volume" effects (i.e., size is underestimated because of the distance of the cut from the longest diameter; such a lesion may appear to have responded or progressed on subsequent examinations, when, in fact, they remain the same size [Fig. 1]). This minimum lesion size for a given slice thickness at baseline ensures that any lesion appearing smaller on subsequent examinations will truly be decreasing in size. The longest diameter of each target lesion should be selected in the axial plane only.

The type of CT scanner is important regarding the slice thickness and minimum-sized lesion. For spiral (helical) CT scanners, the minimum size of any given lesion at baseline may be 10 mm, provided the images are reconstructed contiguously at 5-mm intervals. For conventional CT scanners, the minimum-sized lesion should be 20 mm by use of a contiguous slice thickness of 10 mm.

The fundamental difference between spiral and conventional CT is that conventional CT acquires the information only for the particular slice thickness scanned, which is then expressed as a two-dimensional representation of that thickness or volume as a gray scale image. The next slice thickness needs to be scanned before it can be imaged and so on. Spiral CT acquires the data for the whole volume imaged, typically the whole of the thorax or upper abdomen in a single breath hold of about 20–30 seconds. To view the images, a suitable reconstruction algorithm is selected, by the machine, so the data are appropriately imaged. As suggested above, for spiral CT, 5-mm reconstructions can be made, thereby allowing a minimum-sized lesion of 10 mm.

Spiral CT is now the standard in most hospitals involved in cancer

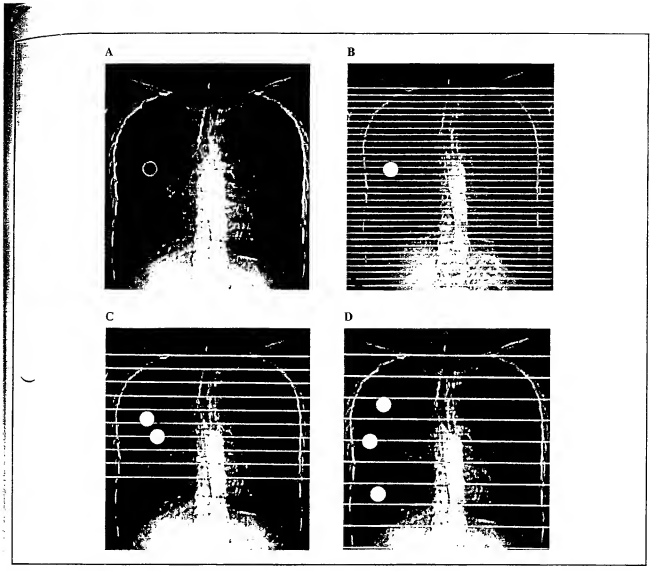

Fig 1. A) Computed tomography (CT) "scannogram" of the thorax with a simulated 20-mm lesion in the right mid-zone. B) CT "scannogram" of the thorax with contiguous slices of 10-mm thickness. Each volume within the slice thickness is scanned, and the average attenuation coefficient (i.e., density of multiple small cubes [voxels]) is represented spatially in two dimensions (pixels) as a cross-sectional image on a gray scale. It is important to note each line on the figure is a spatial representation of the average density for the structures that pass through that slice thickness, and the line does not represent a thin "cut" through it at that level. Therefore, a lesion of at least 20 mm will appear about its true diameter on at least one image because sufficient volume of the lesion is present so as not to average it down substantially. C) CT scannogram performed at 15-mm intervals. Depending on how much of the tumor is within the slice thickness, the average density may be substantially underestimated, as in the upper of the two lesions, or it may approximate the true tumor diameter, lower lesion. This is an oversimplification of the process but illustrates the point without going into the physics of CT reconstruction. D) CT scannogram performed at 24-mm intervals and of 10-mm thickness. The lesion may be imaged through its diameter, it may be partially imaged, or it may not be imaged at all. This is the equivalent of imaging a very small lesion and trying to determine whether its true diameter has changed from one examination to the next.

management in the United States, Europe, and Japan, so the above comments related to spiral CT are pertinent. However, some institutions involved in clinical trials will have conventional CT, but the number of these scanners will decline as they are replaced by spiral CT.

Other body parts, where CT scans are of different slice thickness (such as the neck, which is typically 5-mm thickness), or in the young pediatric population, where the slice thickness may be different, the minimum-sized lesion allowable for measurability of the lesion may be different. However, it should be double the slice thickness. The slice thickness and the minimum-sized lesion should be specified in the study protocol.

In patients in whom the abdomen and pelvis have been imaged, oral contrast agents should be given to accentuate the bowel against other soft-tissue masses. This procedure is almost universally undertaken on a routine basis.

Intravenous contrast agents should also be given, unless contraindicated for medical reasons such as allergy. This is to accentuate vascular structures from adjacent lymph node masses and to help enhance liver and other visceral metastases. Although, in clinical practice, its use may add little, in the context of a clinical study where objective response rate based on measurable disease is the end point, unless an intravenous contrast agent is given, a substantial number of otherwise measurable lesions will not be measurable. The use of intravenous contrast agents may sometimes seem unnecessary to monitor the evolution of specific disease sites (e.g., in patients in whom the disease is apparently restricted to the periphery of the lungs). However, the aim of a clinical

study is to ensure that lesions are truly resolving, and there is no evidence of new disease at other sites scanned (e.g., small metastases in the liver) that may be more easily demonstrated with the use of intravenous contrast agent that should, therefore, also be considered in this context.

The method of administration of intravenous contrast agents is variable. Rather than try to institute rigid rules regarding methods for administering contrast agents and the volume injected, it is appropriate to suggest that an adequate volume of a suitable contrast agent should be given so that the metastases are demonstrated to best effect and a consistent method is used on subsequent examinations for any given patient.

All images from each examination should be included and not "selected" images of the apparent lesion. This distinction is intended to ensure that, if a review is undertaken, the reviewer can satisfy himself/herself that no other abnormalities coexist. All window settings should be included, particularly in the thorax, where the lung and soft-tissue windows should be considered.

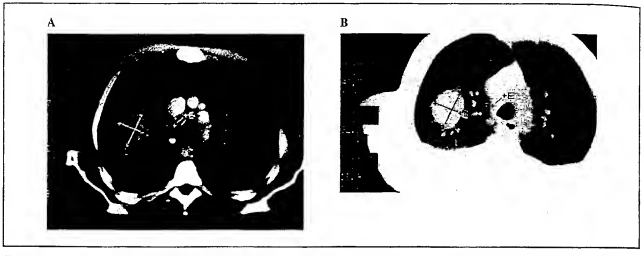

Fig 2. A) Computed tomography (CT) scan of the thorax at the level of the carina on "soft-tissue" windows. Two lesions have been measured with calipers. The intraparenchymal lesion has been measured bidimensionally, using the greatest diameter and the greatest perpendicular distance. Unidimensional measurements require only the greatest diameter to be measured. The anterior–carinal lymph node has been measured using unidimensional criteria. B) The same image as above imaged on "lung" windows, with the calipers remaining as they were for the soft-tissue measurements. The size of the lung lesion appears different. The anterior–carinal lymph node cannot be measured on these windows. The same windows should be used on subsequent examinations to measure any lesions. Some favor soft-tissue windows, so paratracheal, anterior, and subcarinal lesions may be followed on the same settings as intraparenchymal lesions.

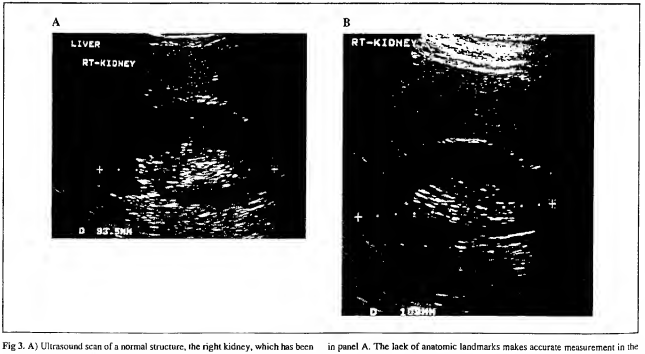

Fig 3. A) Ultrasound scan of a normal structure, the right kidney, which has been measured as 93 mm with the use of calipers. B) Ultrasound scan of the same kidney taken a few minutes later when it measures 108 mm. It appears to have increased in size by 16%. The difference is due to foreshortening of the kidney in panel A. The lack of anatomic landmarks makes accurate measurement in the same plane on subsequent examinations difficult. One has to hope that the measurements given on the hard copy film are a true and accurate reflection of events.

Lesions should be measured on the same window setting on each examination. It is not acceptable to measure a lesion on lung windows on one examination and on soft-tissue settings on the next (Fig. 2). In the lung, it does not really matter whether lung or soft-tissue windows are used for intraparenchymal lesions, provided a thorough assessment of nodal and parenchymal disease has been undertaken and the target lesions are measured as appropriate by use of the same window settings for repeated examinations throughout the study.

Use of MRI is a complex issue. MRI is entirely acceptable and capable of providing images in different anatomic planes. It is, therefore, important that, when MRI is used, lesions must be measured in the same anatomic plane by use of the same imaging sequences on subsequent examinations. MRI scanners vary in the images produced. Some of the factors involved include the magnet strength (high-field magnets require shorter scan times, typically 2–5 minutes), the coil design, and patient cooperation. Wherever possible, the same scanner should be used. For instance, the images provided by a 1.5-Tesla scanner will differ from those provided by a 0.5-Tesla scanner. Although comparisons can be made between images from different scanners, such comparisons are not ideal. Moreover, many patients with advanced malignancy are in pain, so their ability to remain still for the duration of a scan sequence—on the order of 2–5 minutes—is limited. Any movement during the scan time leads to motion artifacts and degradation of image quality, so that the examination will probably be useless. For these reasons, CT is, at this point in time, the imaging modality of choice.

Ultrasound examinations should not be used in clinical trials to measure tumor regression or progression of lesions that are not superficial because the examination is necessarily subjective. Entire examinations cannot be reproduced for independent review at a later date, and it must be assumed, whether or not it is the case, that the hard-copy films available represent a true and accurate reflection of events (Fig. 3). Furthermore, if, for example, the only measurable lesion is in the para-aortic region of the abdomen and if gas in the bowel overlies the lesion, the lesion will not be detected because the ultrasound beam cannot penetrate the gas. Accordingly, the disease staging (or restaging for treatment evaluation) for this patient will not be accurate.

The same imaging modality must be used throughout the study to measure disease. Different imaging techniques have differing sensitivities, so any given lesion may have different dimensions at any given time if measured with different modalities. It is, therefore, not acceptable to interchange different modalities throughout a trial and use these measurements. It must be the same technique throughout.

It is desirable to try to standardize the imaging modalities without adding undue constraints so that patients are not unnecessarily excluded from clinical trials.

APPENDIX II. RELATIONSHIP BETWEEN CHANGE IN DIAMETER, PRODUCT, AND VOLUME

Appendix II, Table 2. Relationship between change in diameter, product, and volume*

	Diameter, $2r$	Product, $(2r)^2$	Volume, $4/3\pi r^3$
Response	Decrease	Decrease	Decrease
	50%	75%	87%
Disease progression	Increase	Increase	Increase
	12%	25%	40%
	20%	44%	73%
	25%	56%	95%
	30%	69%	120%

*Shaded areas represent the response evaluation criteria in solid tumors (diameter) and World Health Organization (product) criteria for change in tumor size to meet response and disease progression definitions.

APPENDIX III. RESPONSE EVALUATION CRITERIA IN SOLID TUMORS (RECIST) WORKING GROUP AND SPECIAL ACKNOWLEDGMENTS

RECIST Working Group

P. Therasse (Chair), J. Verweij, M. Van Glabbeke, A. T. van Oosterom, European Organization for Research and Treatment of Cancer (Brussels, Belgium); S. G. Arbuck, R. S. Kaplan, M. C. Christian, National Cancer Institute, United States (Bethesda, MD); E. Eisenhauer, National Cancer Institute of Canada Clinical Trials Group (Kingston); S. Gwyther, East Surrey Hospital (Redhill, U.K.); and J. Wanders, New Drug Development Office Oncology (Amsterdam, The Netherlands).

Retrospective Analyses

L. A. Rubinstein, National Cancer Institute, United States; B. K. James, A. Muldal, W. Walsh, National Cancer Institute of Canada Clinical Trials Group; S. Green, Southwest Oncology Group (Seattle, WA); M. Terenziani, National Cancer Institute (Milan, Italy); D. Vena, Emmes Corporation (Rockville, MD); R. Canetta, J. Burroughs, Bristol-Myers Squibb (Wallingford, CT); A. Riva, M. Murawsky, Rhone-Poulenc Rorer Pharmaceuticals Inc. (Paris, France).

APPENDIX IV. PARTICIPANTS IN THE OCTOBER 1998 WORKSHOP TO DEVELOP THE FINAL RESPONSE EVALUATION CRITERIA IN SOLID TUMORS (RECIST) DOCUMENT AND FURTHER ACKNOWLEDGMENTS

Participants

S. C. S. Kao, Children's Cancer Study Group (Iowa City, IA); D. Grinblatt, Cancer and Leukemia Group B (CALGB) (Chicago, IL); B. Giantonio, Eastern Cooperative Oncology Group (ECOG) (Philadelphia, PA); F. B. Stehman, Gynecologic Oncology Group (GOG) (Indianapolis, IN); A. Trotti, Radiation Therapy Oncology Group (Tampa, FL); C. A. Coltman, Southwest Oncology Group (SWOG) (San Antonio, TX); R. E. Smith, National Surgical Adjuvant Breast and Bowel Project (Pittsburgh, PA); J. Zalcberg, Peter MacCallum Cancer Institute (Melbourne), Australia; N. Saijo, National Cancer Center Hospital (Tokyo, Japan); Y. Fujiwara, National Institute of Health Sciences (Tokyo); G. Schwartzmann, Hospital de Clinicas de Porto Alegre (Brazil); A. Klein, Health Canada, Bureau of Pharmaceutical Assessment (Ottawa, ON); B. Weinerman, National Cancer Institute of Canada Clinical Trials Group (Kingston, ON); D. Warr, Ontario Cancer Institute/Princess Margaret Hospital (Toronto); P. Liati, South Europe New Drugs Organization (Milan, Italy); S. Einstein, Bio-Imaging Technologies (West Trenton, NJ); S. Négrier, L. Ollivier, Fédération Nationale des Centres de Lutte contre le Cancer (Paris, France); M. Marty, International Cancer Cooperative Group/French Drug Agency (Paris); H. Anderson, A. R. Hanauske, European Organization for Research and Treatment of Cancer (EORTC) (Brussels, Belgium); M. R. Mirza, Odense University Hospital (Denmark), J. Ersboll, The European Agency for the Evaluation of Medicinal Products (Bronshoj, Denmark), C. Pagonis, Cancer Research Campaign (London, U.K.); S. Hatty, Eli Lilly and Co., (Surrey, U.K.); A. Riva, Rhone-Poulenc Rorer Pharmaceuticals Inc. (Paris); C. Royce, GlaxoWellcome (Middlesex, U.K.); G. Burke, Novartis Pharma AG (Basel, Switzerland); I. Horak, Janssen Research Foundation (Beerse, Belgium); G. Hoctin-Boes, Zeneca (Macclesfield Cheshire, U.K.); C. Weil, Bristol-Myers Squibb (Waterloo, Belgium); M. G. Zurlo, Pharmacia & Upjohn (Milan), S. Z. Fields, SmithKline Beecham Pharmaceuticals (Collegeville, PA), B. Osterwalder, Hoffmann-La Roche Inc. (Basel); Y. Shimamura, Taiho Pharmaceutical Co. Ltd. (Tokyo); and M. Okabe, Kyowa-Hakko-Kogyo Co. Ltd (Tokyo).

Additional comments were received from the following:

A. Hamilton, R. De Wit, E. Van Cutsem, J. Wils, J.-L. Lefèbvre, I. Vergote, M. S. Aapro, J.-F. Bosset, M. Hernandez-Bronchud, D. Lacombe, H. J. Schmoll, E. Van Limbergen, P. Fumoleau, A. Bowman, U. Bruntsch, EORTC (Brussels); B. Escudier, P. Thiesse, N. Tournemaine, P. Troufleau, C. Lasset, F. Gomez, Fédération Nationale des Centres de Lutte contre le Cancer (Paris); G. Rustin, Mount Vernon Hospital (Northwood Middlesex, U.K.); S. B. Kaye, Western Infirmary (Glasgow, U.K.); A. Goldhirsch, F. Nolè, G. Zampino, F. De Braud, M. Colleoni, E. Munzone, T. De Pas, International Breast Cancer Study Group and Istituto Europeo di Oncologia (Milan); M. Castiglione, J. F. Delaloye, A. Roth, C. Sessa, D. Hess, B. Thürlimann, C. Böhme, T. Cerny, U. Hess, Schweizer Arbeitsgemeinschaft für Klinische Krebsforschung (Bern, Switzerland); H. J. Stewart, Scottish Cancer Therapy Network (Edinburgh, U.K.); A. Howell, J. F. R. Robertson, United Kingdom Coordinating Committee on Cancer Research (Nottingham); K. Noever, Bio-Imaging Technologies (Monheim, Germany); M. Kurihara, Toyosu Hospital, SHOWA University (Tokyo); L. Seymour, J. Pater, J. Rusthoven, F. Shepherd, J. Maroun, G. Cairncross, D. Stewart, K. Pritchard, National Cancer Institute of Canada Clinical Trials Group (Kingston); T. Uscinowicz, Health Canada, Bureau of Pharmaceutical Assessment (Ottawa); I. Tannock, Princess Margaret Hospital (Toronto); M. Azab, QLT Phototherapeutics (Vancouver, Canada); V. H. C. Bramwell, Canadian Sarcoma Group (London); P. O'Dwyer, ECOG (Philadelphia); A. Martin, S. Ellenberg, U.S. Food and Drug Administration (Rockville, MD); C. Chow, D. Sullivan, A. Murgo, A. Dwyer, J. Tatum, National Cancer Institute (Bethesda, MD); R. Schilsky, CALGB (Chicago, IL); J. Crowley, S. Green, SWOG (Seattle, WA); R. Park, GOG (Philadelphia, PA); V. Land, B. D. Fletcher, Pediatric Oncology Group (Chicago, IL); B. Hillman, University of Virginia (Charlottesville); F. Muggia, New York University Medical Center (New York); C. Erlichman, Mayo Clinic (Rochester, MN); L. H. Schwartz, Memorial Sloan-Kettering Cancer Center (New York, NY); S. P. Balcerzak, Ohio State University Health Sciences Center (Columbus); G. Fleming, CALGB (Chicago); G. Sorensen, Harvard University (Cambridge, MA); H. Levy, Thomas Jefferson University (Philadelphia); N. Patz, Duke University (Durham, NC); C. Visseren-Grul, Eli Lilly Nederland BV (Nieuwegein, The Netherlands)/J. Walling, Lilly Research Laboratories (Indianapolis); P. Hellemans, Janssen Research Foundation (Beerse, Belgium); L. Finke, Merck (Darmstadt, Germany); A. Man, N. Barbet, Novartis Pharma AG (Basel); G. Massimini, Pharmacia & Upjohn (Milan); J. Jimeno, Pharma Mar (Madrid, Spain); I. Hudson, SmithKline Beecham Pharmaceuticals (Essex, U.K.); and J. Krebs, R. A. Beckman, S. Lane, D. Fitts, SmithKline Beecham Pharmaceuticals (Collegeville).

Appendix V. Retrospective Comparison of Response/Disease Progression Rates Obtained With the World Health Organization (WHO)/Southwest Oncology Group Criteria and the New Response Evaluation Criteria in Solid Tumors (RECIST) Criteria

To evaluate the hypothesis by which unidimensional measurement of tumor lesions may substitute for the usual bidimensional approach, a number of retrospective analyses have been undertaken. The results of these analysis are given below in this section.

1. Comparison of Response and Disease Progression Rates by Use of WHO (or Modified WHO) or RECIST Methods

1.1. Trials Evaluated

No specific selection criteria were employed except that trial data had to include serial (repeated) records of tumor measurements. Several groups evaluated their own data on one or more such studies (National Institute of Canada Clinical Trials Group, Kingston, ON; U.S. National Cancer Institute, Bethesda, MD; and Rhone-Poulenc Rorer Pharmaceuticals Inc., Paris, France) or made data available for evaluation to the U.S. National Cancer Institute (Southwest Oncology Group and Bristol-Myers Squibb, Wallingford, CT).

1.2. Response Criteria Evaluated

Not all databases were assessed for all response outcomes. At the outset of this process, the most interest was in the assessment of complete plus partial response rate comparisons by both the WHO and new RECIST criteria. Once these data suggested no impact of using the new criteria on the response rate, several more databases were analyzed for the impact of the use of the new criteria not only on complete response plus partial response but also on stable disease and progressive disease rates (*see* Appendix V, Table 4) and on time to disease progression (*see* Appendix V, Table 5).

1.3. Methods of Comparison

For each patient in each study, baseline sums were calculated (sum of products of the two longest diameters in perpendicular dimensions for WHO and sum of longest diameters for RECIST). After each assessment, when new tumor measures were available, the sums were recalculated. Patients were assigned complete response, partial response, stable disease, and progressive disease as their "best" response on the basis of achieving the measurement criteria as indicated in Appendix V, Table 3. For both WHO and RECIST, a minimum interval of 4 weeks was required to consider complete response and partial response confirmed. Each patient could, therefore, be assigned a best response according to each of the two criteria. The overall response and disease progression rates could be calculated for the population studied for each trial or dataset examined.

(*Note:* For WHO progressive disease, as is the convention in most groups, an increase in sums of products was required, not an increase in only one lesion.)

1.4. Results

2. Evaluation of Time to Disease Progression

Time to disease progression was evaluated, comparing WHO criteria with RECIST in a dataset provided by the Southwest Oncology Group

Appendix V, Table 3. Definition of best response according to WHO or RECIST criteria*

Best response	WHO change in sum of products	RECIST change in sums longest diameters
CR	Disappearance; confirmed at 4 wks†	Disappearance; confirmed at 4 wks†
PR	50% decrease; confirmed at 4 wks†	30% decrease; confirmed at 4 wks†
SD	Neither PR nor PD criteria met	Neither PR nor PD criteria met
PD	25% increase; no CR, PR, or SD documented before increased disease	20% increase; no CR, PR, or SD documented before increased disease

*WHO = World Health Organization; RECIST = Response Evaluation Criteria in Solid Tumors; CR = complete response, PR = partial response, SD = stable disease, and PD = progressive disease.

†For the Bristol-Myers Squibb (Wallingford, CT) dataset, only unconfirmed CR and PR have been used to compare best response measured in one dimension (RECIST criteria) versus best response measured in two dimensions (WHO criteria). The computer flag identifying confirmed response in this dataset could not be used in the comparison for technical reasons.

Appendix V. Table 4. Comparison of RECIST (unidimensional) and WHO (bidimensional) criteria in the same patients recruited in 14 different trials*

Tumor site/type	Criteria	No. of patients evaluated	Best response				RR	PD rate
			CR	PR	SD	PD		
Breast†	WHO	48	4	22			54%	
	RECIST	48	4	22			54%	
Breast‡	WHO	172	4	36			23%	
	RECIST	172	4	40			26%	
Brain†	WHO	31	12	10			71%	
	RECIST	31	12	10			71%	
Melanoma†	WHO	190	9	37			24%	
	RECIST	190	9	34			23%	
Breast§	WHO	531	50	102			29%	
	RECIST	531	50	108			30%	
Colon§	WHO	1096	12	137			14%	
	RECIST	1096	12	133			13%	
Lung§	WHO	1197	60	317			32%	
	RECIST	1197	60	318			32%	
Ovary§	WHO	554	24	108			24%	
	RECIST	554	24	105			23%	
Lung†	WHO	24	0	4	16	4	17%	17%
	RECIST	24	0	4	19	1	17%	4%
Colon†	WHO	31	1	6	15	9	23%	29%
	RECIST	31	1	5	16	9	21%	29%
Sarcoma†	WHO	28	1	4	13	10	18%	36%
	RECIST	28	1	5	17	5	21%	18%
Ovary†	WHO	45	0	7	19	19	16%	42%
	RECIST	45	0	6	21	18	13%	40%
Breast‖	WHO	306	18	114	117	57	43%	19%
	RECIST	306	18	108	124	56	41%	18%
Breast‖	WHO	360	10	73	135	142	23%	39%
	RECIST	361	10	70	139	142	22%	39%
Total (all studies where tumor response was evaluated)	WHO	4613	205	977			25.6%	
	RECIST	4614	205	968			25.4%	
Total (all studies where PD as well as CR + PR were evaluated)	WHO	794			315	241		30.3%
	RECIST	795			336	231		29%

*WHO = World Health Organization *(3)*; RECIST = Response Evaluation Criteria in Solid Tumors; CR = complete response; PR = partial response; SD = stable disease; PD = progressive disease; and RR = response rate.

†Data from the National Cancer Institute of Canada Clinical Trials Group phase II and III trials.

‡Data from the National Cancer Institute, United States phase III trial.

§Data from Bristol-Myers Squibb (Wallingford, CT) phase II and III trials.

‖Data from Rhone-Poulenc Rorer Pharmaceuticals Inc., (Paris, France) phase III trials (*note:* one patient in this database had unidimensional measured lesions only ... could not be evaluated with the WHO criteria).

Appendix V, Table 5. Proportions of patients with disease progression by different assessment methods*

	No. of patients	%
Total No. of progressors	234	100
Progress by appearance of new lesions†	118	50
Progress by increase in pre-existing measurable disease	116	50
Same date of disease progression by WHO and RECIST criteria	215	91.9
Different date of disease progression	19	8.1
Earlier PD with WHO criterion	17	7.3
Earlier PD with unidimensional criterion	2	0.9

*PD = progressive disease; WHO = World Health Organizaiton; and RECIST = Response Evaluation Criteria in Solid Tumors.

†Also includes a few patients with PD because of marked increase of nonmeasurable disease.

Appendix V, Table 6. Magnitude of time to disease progression disagreements when differences existed*

	No. of patients	% (of 234, see above)
No. of progressors with differing progression dates	19	8.1
8–9 wks' difference	3	1.3
12 wks' difference	1	0.4
24–31 wks' difference†	2	0.9
Difference uncertain due to censoring of either WHO or RECIST progression time‡	13	5.6

*WHO = World Health Organization; RECIST = Response Evaluation Criteria in Solid Tumors.

†For one patient, progression by RECIST (one-dimension) criteria preceded that by WHO criteria by 24 weeks due primarily to one-dimensional growth. For a second patient, with a colon tumor that increased in cross-section by 25%, then regressed completely, and then recurred, progression by WHO criteria preceded that by RECIST criteria by 31 weeks.

‡As indicated in Appendix V, Table 6, 13 of the 19 patients had uncertain disease progression time differences when comparing RECIST and WHO criteria. In these patients, the RECIST progression criteria were not met by the time that disease progression by Southwest Oncology Group (SWOG) criteria *(5)* had occurred (50% increase or a 10 cm^2 increase in tumor cross-section). Notably, six of these patients had the same disease progression dates determined by use of WHO (25% bidimensional increase) and SWOG (50% bidimensional increase) criteria. Since 20% unidimensional increase (RECIST) is equivalent to approximately 44% bidimensional increase, it is likely, although not certain, that disease progression by RECIST unidimensional criteria would have occurred soon after disease progression by SWOG and WHO criteria. For three patients, the difference between the WHO and SWOG 50% bidimensional increase was 10–12 weeks. Again, it is likely, although it cannot be proven, that RECIST criteria would have been met soon after. The remaining four of the 13 patients where difference between WHO and RECIST progression times are uncertain were categorized as progressive disease following SWOG's criteria *(5)* because of an increase of the tumor surface of greater than or equal to 10 cm^2. For these patients, the magnitude of the difference is entirely uncertain.

(SWOG). Since SWOG criteria *(5)* for disease progression is a 50% increase in the sum of the products, or new disease, or an absolute increase of 10 cm^2 in the sum of the products, this dataset provided the means of assessing the impact of time to disease progression differences between a 25% increase in the sum of the products and a 20% increase in the sum of the longest diameters (equivalent to approximately a 44% increase in the product sum).

2.1. Dataset Evaluated

The dataset includes 234 patients with progressive disease as defined by the SWOG *(5)*. All patients had baseline measurable disease followed by the same technique(s) until disease progression. The tumor types included were melanoma and colorectal, lung, and breast cancers.

References

(1) Zubrod CG, Schneiderman SM, Frei E III, Brindley C, Gold GL, Schnider B, et al. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and thio phosphoamide. J Chronic Dis 1960;11:7–33.

(2) Gehan E, Schneidermann M. Historical and methodological developments in clinical trials at the National Cancer Institute. Stat Med 1990;9: 871–80.

(3) WHO handbook for reporting results of cancer treatment. Geneva (Switzerland): World Health Organization Offset Publication No. 48; 1979.

(4) Miller AB, Hogestraeten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47:207–14.

(5) Green S, Weiss GR. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. Invest New Drugs 1992;10:239–53.

(6) James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A, et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. J Natl Cancer Inst 1999;91:523–8.

Note

Manuscript received June 24, 1999; revised November 30, 1999; accepted December 3, 1999.

AHadm+5

ARTICLES

3228
22.30

New Colorimetric Cytotoxicity Assay for Anticancer-Drug Screening

Philip Skehan, Ritsa Storeng, Dominic Scudiero, Anne Monks, James McMahon, David Vistica, Jonathan T. Warren, Heidi Bokesch, Susan Kenney, Michael R. Boyd*

We have developed a rapid, sensitive, and inexpensive method for measuring the cellular protein content of adherent and suspension cultures in 96-well microtiter plates. The method is suitable for ordinary laboratory purposes and for very large-scale applications, such as the National Cancer Institute's disease-oriented in vitro anticancer-drug discovery screen, which requires the use of several million culture wells per year. Cultures fixed with trichloroacetic acid were stained for 30 minutes with 0.4% (wt/vol) sulforhodamine B (SRB) dissolved in 1% acetic acid. Unbound dye was removed by four washes with 1% acetic acid, and protein-bound dye was extracted with 10 m*M* unbuffered Tris base [tris (hydroxymethyl)aminomethane] for determination of optical density in a computer-interfaced, 96-well microtiter plate reader. The SRB assay results were linear with the number of cells and with values for cellular protein measured by both the Lowry and Bradford assays at densities ranging from sparse subconfluence to multilayered supraconfluence. The signal-to-noise ratio at 564 nm was approximately 1.5 with 1,000 cells per well. The sensitivity of the SRB assay compared favorably with sensitivities of several fluorescence assays and was superior to those of both the Lowry and Bradford assays and to those of 20 other visible dyes. The SRB assay provides a colorimetric end point that is nondestructive, indefinitely stable, and visible to the naked eye. It provides a sensitive measure of drug-induced cytotoxicity, is useful in quantitating clonogenicity, and is well suited to high-volume, automated drug screening. SRB fluoresces strongly with laser excitation at 488 nm and can be measured quantitatively at the single-cell level by static fluorescence cytometry. [J Natl Cancer Inst 82:1107–1112, 1990]

The recent emergence of computer-interfaced fiber-optic readers for 96-well microtiter plates has provided the basis for rapid in vitro cytotoxicity analysis that is particularly well suited to preclinical drug discovery and development. Although a variety of spectrophotometric methods are available for the analysis of large numbers of cells, few possess the sensitivity required by the semi-micro dimensions of microtiter plates. Fewer still are suitable for the very high volume of samples involved in large-scale drug screens, such as the disease-oriented in vitro anticancer-drug discovery project of the National Cancer Institute (NCI). This project tests 10,000 or more samples each year in a manner that requires the analysis of several million individual wells (*1*).

We compared the abilities of 21 histological dyes to measure cell density and drug cytotoxicity in 96-well microtiter plates. The dyes bind electrostatically to macromolecular counterions in cells fixed with trichloroacetic acid (TCA) (*2–4*), which allows their binding and solubilization to be controlled by variations in pH (*2*). In one pH range, the dyes bind stoichiometrically to target macromolecular counterions, whereas in another, they can be quantitatively extracted for measurement of optical density (*2*).

Thirteen of the dyes stained well enough to provide an adequate basis for assay of cytotoxicity in 96-well plates. Optimized protocols were developed for the seven best dyes. Four of these were anionic protein stains with sulfonic or sulfinic groups that bind electrostatically to protein basic amino acid residues under mildly acidic conditions (*2–5*). These dyes can be quantitatively extracted from cells and solubilized for optical density measurement by weak bases (*2*). The other three dyes were cationic dyes that bind electrostatically to macromolecular negative fixed charges (*3,4*). Under mildly basic conditions, these dyes bind to proteins, RNA, DNA, and glycosaminoglycans, serving as gen-

Received September 5, 1989; revised December 11, 1989; accepted March 15, 1990.

P. Skehan, R. Storeng, J. McMahon, D. Vistica, J. T. Warren, S. Kenney (Developmental Therapeutics Program, Division of Cancer Treatment), D. Scudiero, A. Monks, H. Bokesch (Program Resources, Inc.), National Cancer Institute, Frederick Cancer Research Facility, Frederick, MD.

M. R. Boyd, Developmental Therapeutics Program, Division of Cancer Treatment, National Cancer Institute, Bethesda, MD.

**Correspondence to:* Philip Skehan, M.D., Developmental Therapeutics Program, Division of Cancer Treatment, Bldg. 560, Rm. 3260, National Cancer Institute, Frederick Cancer Research Facility, Frederick, MD 21701.

eral biomass stains (*4*). They can be extracted from cells with a weak acid.

Materials and Methods

Cells

We performed preliminary experiments with the human A-2780 ovarian, HT-29 colon, and UO-31 renal tumor cell lines to identify the most promising dyes, which were subsequently examined in detail with some or all of the cell lines currently used in the NCI's in vitro anticancer-drug screen (*6*).

Stock cultures were grown in T-75 flasks containing 50 mL of RPMI-1640 medium with glutamine, bicarbonate, and 5% fetal calf serum. Medium was changed at 48-hour intervals. Cells were dissociated with 0.25% trypsin and 3 m*M* 1,2-cyclohexanediaminetetraacetic acid in NKT buffer (137 m*M* NaCl, 5.4 m*M* KCl, and 10 m*M* Tris; pH 7.4). Experimental cultures were plated in microtiter plates (Costar, Cambridge, MA) containing 0.2 mL of growth medium per well at densities of 1,000–200,000 cells per well.

Dyes

Dyes were purchased from Sigma Chemical Co., St. Louis, MO. Preliminary studies were conducted with each of these 21 dyes to determine whether each stained cells more intensely at acidic, neutral, or basic pH (*2*). The anionic dyes bromophenol blue, chromotrope 2R, Coomassie brilliant blue, naphthol yellow S, orange G, and sulforhodamine B (SRB) were dissolved in 1% acetic acid for cell staining and extracted from cells with 10 m*M* unbuffered Tris base [tris(hydroxymethyl)aminomethane]. The cationic dyes acridine orange, azure A, azure B, azure C, cresyl violet acetate, methyl green, methylene blue, phenosafranin, safranin O, thionin, and toluidine blue O were dissolved in unbuffered 10 m*M* Tris base to stain cells and were resolubilized for measurement of optical density with either 1% or 10% acetic acid. The cationic dyes ethidium bromide, propidium iodide, and pyronin B were dissolved in water for staining. Although these are excellent fluorescent dyes (*7*), their staining intensity was poor at visible wavelengths. Crystal violet was dissolved in 10% ethanol and 90% water at a neutral pH; its staining intensity varied considerably from one cell line to another. The absorption maximum of each dye in its solubilizing solution was determined with a DU-70 scanning spectrophotometer (Beckman Instruments, Inc., Fullerton, CA).

Cell Fixation

Washing cultures with buffer prior to fixation to remove serum protein commonly caused cell detachment and loss. To avoid this potential problem, cultures were fixed with TCA before washing. Cells attached to the plastic substratum were fixed by gently layering 50 μL of cold 50% TCA (4 °C) on top of the growth medium in each well to produce a final TCA concentration of 10%. The cultures were incubated at 4 °C for 1 hour and then washed five times with tap water to remove TCA, growth medium, and low-molecular-weight metabolites, and serum protein. Plates were air dried and then stored until use. Background optical densities were measured in wells incubated with growth medium without cells.

Cells in suspension were allowed to settle out of solution. When these cells were physically resting on the bottom of the wells, 50 μL of cold 80% TCA (4 °C) was gently layered on top of the overlying growth medium. The cultures were left undisturbed for 5 minutes and then refrigerated at 4 °C for an additional hour of fixation. This procedure led to the attachment of single cell suspensions to the plastic substratum, provided that cells were in contact with it when the fixative was applied. This method was as effective in promoting cell attachment as were cytospinning and using the macromolecular adhesive Cell-Tak (Biopolymers, Farmington, CT). However, it did not adequately attach cells that grew as floating aggregates rather than as single cell suspensions. Small cell lung carcinoma lines were particularly unsuited to this method of fixation.

Following fixation, suspension cultures were processed with procedures identical to those used for cultures of cells attached to the plastic substratum. After cells were stained and washed, individual wells were checked for cell detachment (clear spots or regions in the normally homogeneous pink carpet of cells), a potentially important source of artifact with cell suspensions. Although 80% TCA caused cells from suspensions to adhere to the plastic substratum, this attachment was extremely sensitive to movement, and very gentle handling of both the cells and the TCA was required. The efficiency of this attachment varied with cell type: cells from some cell lines were well attached by this method, while others were not.

Organic solvents such as ethanol and methanol were not suitable fixatives for the dye assays. When mixed with growth medium, these solvents generated intense interfacial shearing forces, which could be seen by phase-contrast microscopy to rip cells from the substratum, lysing many in the process. These shearing forces represented a major source of fixation artifacts and were not diminished by prior aspiration of the growth medium. Aqueous fixatives did not produce this effect. TCA and perchloric acid both gave extremely rapid fixation, and no morphological artifacts were observable by phase-contrast microscopy. Formaldehyde was less satisfactory. It caused the formation of extensive plasma membrane blebs with a concomitant loss of cytoplasmic protein. Glutaraldehyde was unsuitable for the purposes of this study, because of its ability to interfere with dye–protein interactions by reacting with and masking the positive fixed charges of protein amino groups (*8*). In addition, formaldehyde also caused the loss of nuclear structure in some cell lines.

Background Levels

Background levels of SRB staining were sensitive to the length of TCA fixation and serum concentration. Cultures could be left in TCA for several hours with little increase in background optical density (OD), which was typically about 0.035 OD units at 520 nm for 96-well plates. When cells were left in TCA overnight, the background OD doubled. Similarly, the background OD of medium containing 10% fetal calf serum was twice that of medium containing 5% fetal calf serum.

Optimized Staining Protocols

An optimized protocol (*2*) was developed for each of the dyes examined in detail. A plateau staining time was determined from the binding kinetics of a 2% dye solution. Optimal destaining was

achieved by determining the number of washes necessary to remove unbound dye without desorbing cell-associated stain. A supramaximal dye concentration that fully saturated cellular binding sites was determined by dose–response analysis of heavily confluent multilayers.

SRB Assay

TCA-fixed cells were stained for 30 minutes with 0.4% (wt/vol) SRB dissolved in 1% acetic acid. At the end of the staining period, SRB was removed and cultures were quickly rinsed four times with 1% acetic acid to remove unbound dye. The acetic acid was poured directly into the culture wells from a beaker. This procedure permitted rinsing to be performed quickly so that desorption of protein-bound dye did not occur. Residual wash solution was removed by sharply flicking plates over a sink, which ensured the complete removal of rinsing solution. Because of the strong capillary action in 96-well plates, draining by gravity alone often failed to remove the rinse solution when plates were simply inverted. After being rinsed, the cultures were air dried until no standing moisture was visible. Bound dye was solubilized with 10 m*M* unbuffered Tris base (pH 10.5) for 5 minutes on a gyratory shaker.

OD was read in either a UVmax microtiter plate reader (Molecular Devices, Menlo Park, CA) or a Beckman DU-70 spectrophotometer. For maximum sensitivity, OD was measured at 564 nm. Because readings were linear with dye concentrations only below 1.8 OD units, however, suboptimal wavelengths were generally used, so that all samples in an experiment remained within the linear OD range. With most cell lines, wavelengths of approximately 490–530 nm worked well for this purpose.

Optical Density Linearity

Curves of OD versus dye concentration were generally linear to 1.5–2.0 OD units. When the linearity range was exceeded, it was necessary either to dilute an aliquot and reread its OD or to use a suboptimal wavelength as a filter to reduce OD and extend the working range of dye concentrations that fell within the limits of linearity. This second method was generally more convenient but had the disadvantage of reducing resolution at low cell density. The problem was averted by reading samples at two separate wavelengths, then converting one to another with a least-squares linear regression equation determined over the range of OD values that were linear for both wavelengths.

Culture Cell Protein Analysis

Cell protein was measured by the Oyama and Eagle modification of the Lowry method, with bovine serum albumin used as a standard (*9*). The contents of individual wells were digested with 0.5 *M* NaOH. Aliquots of the digest were diluted with 0.5 *M* NaOH to a final volume of 0.4 mL and mixed with 2 mL of Lowry C solution. To this mixture was added 0.2 mL of Folin–Ciocalteu's phenol reagent (Sigma Chemical Co.) diluted 5:4 with distilled water. Color was allowed to develop for 30 minutes, and OD was measured at 660 nm.

Cell protein was also measured by the Bradford Coomassie brilliant blue dye method (*10*) using Pierce protein assay reagent (Pierce Chemical Co., Rockford, IL). The contents of individual wells were digested with 0.1 mL of 0.5 *M* NaOH. The digest was mixed with 4 volumes of 0.5 *M* NaOH and 5 volumes of Pierce reagent and agitated on a gyratory shaker for 5 minutes. Absorbance was then measured at 595 nm. A calibration curve was constructed, with bovine serum albumin used as a standard.

Results

Comparison of Dyes

Of the 21 dyes tested, 13 stained TCA-fixed cultures sufficiently well to provide the basis for a quantitative assay of cell number and drug cytotoxicity in a 96-well plate. These dyes were acridine orange, azure A, azure B, bromophenol blue, chromotrope 2R, cresyl violet acetate, methylene blue, orange G, phenosafranin, safranin O, SRB, thionin, and toluidine blue O. Four of these dyes were protein stains, while the remainder were general macromolecular biomass stains (*2–5*). The other eight dyes either stained too lightly to be useful or stained different cell lines with widely varying intensity. These dyes were azure C, Coomassie brilliant blue, crystal violet, ethidium bromide, methyl green, naphthol yellow S, propidium iodide, and pyronin B.

The most intensely staining dyes were bromophenol blue and SRB, both of which are protein stains. They were closely followed by thionin, azure A, and toluidine blue O, which are thiazin quinone–imine cationic biomass stains.

There was no clear advantage to any one of these dyes at high cell densities. All commonly produced OD values for confluent cultures that exceeded their linearity limits. At low cell densities, however, SRB was distinctly superior in signal-to-noise ratio (table 1). Results were quantitative at densities above 2,500 cells

Table 1. Optimal staining protocols for selected dyes*

Dye	Concentration (%)	Staining solution†	Minimum stain time (min)	No. of washes	Solubilizing solution†	Optimal wavelength (nm)	Signal-to-noise ratio at 5,000 cells per well	Resolution ($\times 10^3$ cells per well)
SRB	0.4	AcOH	15	4	Tris	564	4.83	1–2
ORG	1.5	AcOH	5	3	Tris	478	2.46	2-3
BPB	1.0	AcOH	20	4	Tris	590	1.92	3–4
CTR	0.25	AcOH	30	3	Tris	508	2.03	4–5
TNN	0.3	Tris	10	4	AcOH	594	1.56	>5
AZRA	0.25	Tris	30	4	AcOH	628	1.18	>10
TBO	0.2	Tris	10	3	AcOH	626	1.16	>10

*Ratios were determined for human HT-29 colon adenocarcinoma cells. ORG = orange G, BPB = bromophenol blue, CTR = chromotrope 2R, TNN = thionin, AZRA = azure A, TBO = toluidine blue, AcOH = acetic acid.
†Solutions were either 1% acetic acid or 10 m*M* unbuffered Tris base.

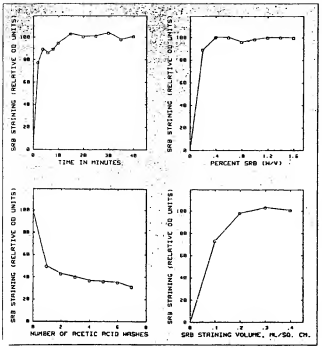

Figure 1. Optimization of SRB assay parameters for HT-29 colon adenocarcinoma cells in 96-well microtiter plates. SRB binding was determined as a function of time (upper left), dye concentration (upper right), number of destaining washes (lower left), and dye volume per unit area of cell culture (lower right). Cells were heavily confluent for optimizing dye volume per unit area and nearly confluent in the other experiments. Optical densities were measured with a UVmax plate reader at the wavelength setting (550, 520, or 490 nm) that provided the greatest sensitivity while remaining below the limit of linearity of 1.8 OD units.

per well and semiquantitative at 1,000 cells per well. Most of the other dyes examined had limits of resolution of 5,000–10,000 cells per well.

As a group, the protein stains tended to provide slightly better resolution than the biomass stains. Thus, two of the organosulfonic protein stains, chromotrope 2R and orange G, while not staining as intensely as some of the other dyes, had signal-to-noise ratios that were among the best at low cell densities (table 1). Their lower staining intensity was offset by the fact that they could be measured at their optimal wavelengths. This gave them an effective sensitivity nearly equal to that of SRB and bromophenol blue and superior to that of thionin, all of which had to be measured at suboptimal wavelengths to ensure that their measured OD values were within the range of linearity.

Optimized Protocols

Optimized protocols were developed for several of the dyes that provided better resolution (table 1). Data for SRB are shown in figure 1. SRB optimizations performed for more than 60 human tumor cell lines gave optimized protocols that were essentially identical. However, optimized parameters did change slightly from one commercial lot of SRB to the next. It is therefore advisable that the staining protocol be individually reoptimized for each new lot of dye.

A threefold increase in sensitivity was achieved by measuring SRB fluorometrically. SRB fluoresced strongly with laser excitation at 488 nm and could be quantitated at the single-cell level by static fluorescence cytometry.

SRB Calibration

Linearity of the SRB assay with cell number was evaluated by plating twofold serial dilutions of rapidly adhering cells (human H-23 lung cancer, SK-MEL-28 and UACC-62 melanoma, and SKOV-3 ovarian cancer cell lines). Cultures were fixed with TCA as soon as attachment was completed. The SRB assay was linear with the number of cells at densities ranging from 1% to more than 200% of confluence (fig. 2). Least-squares linear correlation coefficients for the SRB–cell number relationship in the four cell lines were 0.9999, 0.9998, 0.9989, and 0.9998, respectively. For non-drug-treated HT-29 colon adenocarcinoma cells, regression analysis showed an average correlation coefficient of 0.9727 for the SRB and Bradford assays. For 37 drugs producing 135 data points on the decreasing portions of their dose–response curves, the correlation coefficient for the SRB and Lowry assays with HT-29 cells was 0.9855 (fig. 3). All correlations were statistically significant at $P < .001$.

From the best-fit parameters of the least-squares analyses, the cell protein determination by the Lowry assay was equal to

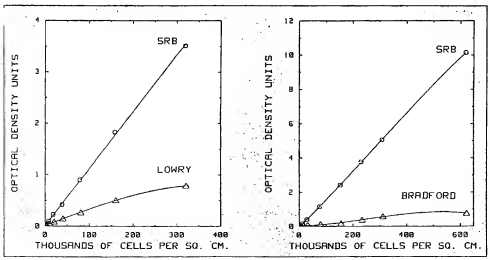

Figure 2. Calibration of the SRB, Lowry, and Bradford assays vs. number of cells. Cells were plated at densities of 25,000–600,000 cells per square centimeter in 96-well microtiter plates with a surface area of 0.32 cm^2. This corresponds to a density range of approximately 1%–200% of confluence. Cultures were fixed with TCA as soon as cells were attached. Growth during this period was negligible.

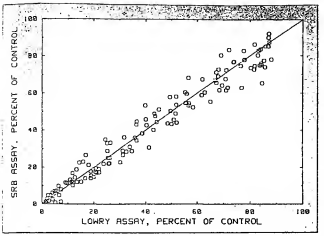

Figure 3. Comparison of the SRB and Lowry assays in evaluating drug-induced cytotoxicity. HT-29 colon adenocarcinoma cells were incubated for 48 hr with six concentrations of each of 37 clinical and experimental anticancer drugs. At the end of the incubation period, replicate cultures were separately evaluted by the SRB and Lowry assays. Each measurement was performed in triplicate. The data represent the 137 points that fell on descending arms of the Lowry dose–response curves. The least-squares correlation coefficient for the SRB–Lowry regression was 0.9855 ($P < .001$). Drugs included doxorubicin, 6-azauridine, colchicine, chromomycin A_3, cytarabine, elliptiicine, erythromycin, fluorouracil, homoharringtonine, mercaptopurine, methotrexate, mitomycin, podophyllotoxin, vinblastine, and vincristine.

$(137.6 \times \text{SRB } OD_{520} \text{ units}) + 1.615$, while the cell protein determination by the Bradford assay was equal to $(7.386 \times \text{SRB } OD_{520} \text{ units}) + 0.549$, in micrograms of bovine serum albumin equivalents. (OD_{520} = OD at 520 nm.)

SRB Assay

The SRB assay provided a rapid and sensitive method for measuring the drug-induced cytotoxicity in both attached and suspension cultures in 96-well microtiter plates. Representative dose–response curves for fluorouracil and cisplatin are shown in fig. 4. SRB staining was also of use in assays of colony formation and colony extinction, permitting colony counts to be compared with the cell protein content of the same cultures (data not shown).

Discussion

SRB is a bright pink aminoxanthene dye with two sulfonic groups (*3*). Its histochemistry is similar to that of related dyes, such as Coomassie brilliant blue, bromophenol blue, and naphthol yellow S, which are used widely as protein stains (*2–5*). Under mildly acidic conditions, SRB binds to protein basic amino acid residues in TCA-fixed cells to provide a sensitive index of cellular protein content that is linear over a cell density range of at least 2 orders of magnitude (fig. 2).

Of the dyes examined in the present study, SRB provided the best combination of staining intensity and signal-to-noise ratio (table 1). Its sensitivity is comparable to the sensitivities of some fluorescent dyes (*11,12*) and superior to those of conventional visible dyes (*2,8,10,13–16;* fig. 2). The 100-fold range of linearity of the SRB assay far exceeds that of the Lowry and Bradford assays, eliminating the need for time-consuming and error-prone dilutions of samples with high-protein contents.

Color development in the SRB assay is rapid, stable, and visible. The OD of SRB can be measured over a broad range of visible wavelengths in either spectrophotometers or 96-well plate readers.

With a properly optimized protocol, SRB staining reaches a true and stable end point that does not have to be measured within any fixed period of time. When air dried, both TCA-fixed and SRB-stained samples can be stored indefinitely without deterioration. Tris-solubilized SRB is also stable for extended periods, provided that evaporation does not occur.

The SRB staining method is nondestructive in the sense that it is not necessary to digest samples. This allows cultures from which dye has been extracted to be restained and saved for future reference. The Tris extraction solution, however, does cause some deterioration in the morphology of samples fixed in 5% or 10% TCA or air dried for short periods of time. This deterioration is accompanied by the solubilization and loss of some cell protein. These effects can be reduced by extending fixation, storing air-dried samples for several weeks prior to Tris extraction, and minimizing the time of sample exposure to Tris.

Although the SRB assay was originally developed for cells attached to a plastic substratum, a variation of the method with an elevated TCA concentration was adequate for a number of cell lines in suspension culture, including the murine P388 lymphoma and the human CCRF-CEM, K562, MOLT-4, HL-60, and RPMI-8226 leukemia lines. This modified method was also useful for cell lines with weakly adherent monolayer cells or with adherent cultures that shed floating cells or small aggregates into the surrounding growth medium.

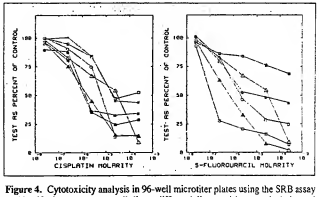

Figure 4. Cytotoxicity analysis in 96-well microtiter plates using the SRB assay to identify human tumor cell lines differentially sensitive to cisplatin and fluorouracil. Cell lines used in the cisplatin experiment were SF-268 central nervous system (CNS) cancer (■), HT-29 colon adenocarcinoma (○), RPMI-8226 leukemia (▼), M19-MEL melanoma (△), H-460 non-small cell lung cancer (▲), OVCAR-4 ovarian cancer (●), and CAKI-I renal cancer (□). Cell lines used in the fluorouracil experiment were XF-498 CNS cancer (■), HCT-116 colon cancer (▲), MOLT-4 leukemia (○), SK-MEL-5 melanoma (△), H-460 non-small cell lung cancer (□), and OVCAR-8 ovarian cancer (●). Cultures were preincubated in growth medium for 24 hr to permit recovery from trypsinization and then incubated for an additional 48 hr with control medium or test solution in growth medium. The H-460, HCT-116, and HT-29 cell lines were plated at 5,000 cells per well; the CAKI-1, M19-MEL, OVCAR-4, OVCAR-8, and SK-MEL-5 cell lines at 10,000 cells per well; the SF-268 cell line at 15,000 cells per well; the RPMI-8226 and XF-498 cell lines at 20,000 cells per well; and the MOLT-4 cell line at 30,000 cells per well.

The SRB assay provides a sensitive method for measuring drug cytotoxicity in culture. In a pilot study of the NCI's in vitro anticancer-drug discovery project, the SRB assay was used to examine the differential sensitivities of 60 human tumor cell lines to more than 1,000 test compounds (*1,6,17,18*). The method appears to offer several advantages over the MTT and XTT assays (*19,20*) for very large-scale drug screening.[1] The SRB assay was simpler, faster, and more sensitive than the MTT assay, provided better linearity with cell number, permitted the use of saturating dye concentrations, was less sensitive to environmental fluctuations, was independent of intermediary metabolism, and provided a fixed end point that did not require a time-sensitive measurement of initial reaction velocity (*21,22*).

References

(1) BOYD MR: Status of the NCI preclinical antitumor drug discovery screen; implications for selection of new agents for clinical trial. *In* Cancer Principles and Practice of Oncology Updates (DeVita VT Jr, Hellman S, Rosenberg SA, eds). Philadelphia: Lippincott, 1990

(2) SKEHAN P, FRIEDMAN SJ: A rapid Naphthol Yellow S method for measuring the cellular protein content of anchorage cultures. In Vitro 21:288–290, 1985

(3) LILLIE RD: H.J. Conn's Biological Stains, 9th ed. Baltimore: Williams & Wilkins, 1977

(4) KIERNAN JA: Histological & Histochemical Methods: Theory & Practice. Oxford: Pergamon Press, 1981

(5) FLORES R: A rapid and reproducible assay for quantitative estimation of proteins using bromophenol blue. Anal Biochem 89:605–611, 1978

(6) STINSON SF, ALLEY MC, KENNEY S, ET AL: Morphologic characterization of human carcinoma cell lines. Proc Am Assoc Cancer Res 30:613, 1989

(7) CRISSMAN HA, MULLANEY PF, STEINKAMP JA: Methods and applications of flow systems for analysis and sorting of mammalian cells. Methods Cell Biol 9:179–246, 1975

(8) SKEHAN P: The mechanisms of glutaraldehyde-fixed sarcoma 180 ascites cell aggregation. J Membrane Biol 24:87–106, 1975

(9) OYAMA VI, EAGLE H: Measurement of cell growth in tissue culture with a phenol reagent (Folin-Ciocalteu). Proc Soc Exp Biol Med 91:305–307, 1956

(10) BRADFORD M: A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principles of protein-dye binding. Anal Biochem 72:248–254, 1976

(11) VAN LAMBALGEN R, LELIEVELD P: The PIT method: An automated in vitro technique for drug toxicity testing. Invest New Drugs 5:161–165, 1987

(12) MCCAFFREY TA, AGARWAL LA, WEKSLER BB: A rapid fluorometric DNA assay for the measurement of cell density and proliferation in vitro. In Vitro 24:247–252, 1988

(13) MOSMANN T: Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. J Immunol Methods 65:55–63, 1983

(14) FINLAY GJ, BAGULEY BL, WILSON WR: A semiautomated microculture method for investigating growth inhibitory effects of cytotoxic compounds on exponentially growing carcinoma cells. Anal Biochem 139:272–277, 1984

(15) LANDEGREN U: Measurement of cell numbers by means of the endogenous enzyme hexosaminidase. Applications to detection of lymphokines and cell surface antigens. J Immunol Methods 67:379–388, 1984

(16) EVERITT E, WOHLFART C: Spectrophotometric quantitation of anchorage-dependent cell number using extraction of Naphthol Blue-Black-stained cellular protein. Anal Biochem 162:122–129, 1987

(17) MONKS A, SCUDIERO D, SKEHAN P, ET AL: Implementation of a pilot scale, high flux anticancer drug screen utilizing disease-oriented panels of human tumor cell lines in culture. Proc Am Assoc Cancer Res 30:607, 1989

(18) SKEHAN P, STORENG R, SCUDIERO D, ET AL: Evaluation of colorimetric biomass stains for assaying in vitro drug effects upon human tumor cell lines. Proc Am Assoc Cancer Res 30:612, 1989

(19) ALLEY MC, SCUDIERO DA, MONKS A, ET AL: Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay. Cancer Res 48:589–601, 1988

(20) SCUDIERO DA, SHOEMAKER RH, PAULL KD, ET AL: Evaluation of a soluble tetrazolium/formazan assay for cell growth and drug sensitivity in culture using human and other tumor cell lines. Cancer Res 48:4827–4833, 1988

(21) VISTICA DT, SKEHAN P, SCUDIERO D, ET AL: Tetrazolium-based assays for cellular viability: A critical examination of parameters which affect formazan production. Proc Am Assoc Cancer Res 30:612, 1989

(22) RUBINSTEIN LV, PAULL KD, SHOEMAKER RH, ET AL: Correlation of screening data generated with a tetrazolium assay (MTT) versus a protein assay (SRB) against a broad panel of human tumor cell lines. Proc Am Assoc Cancer Res 30:607, 1989

[1]MTT = 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide. XTT = 2,3-bis[2-methoxy-4-nitro-5-sulfophenyl]-5-[(phenylamino)carbonyl]-2H-tetrazolium hydroxide.